Negotiating with other minds

de Weerd, Harmen; Verbrugge, Rineke; Verheij, Bart

Published in:
Autonomous Agents and Multi-Agent Systems

DOI:
10.1007/s10458-015-9317-1

Auton Agent Multi-Agent Syst
DOI 10.1007/s10458-015-9317-1

Negotiating with other minds: the role of recursive theory of mind in negotiation with incomplete information

Harmen de Weerd[1] · **Rineke Verbrugge**[1] · **Bart Verheij**[1]

Abstract Theory of mind refers to the ability to reason explicitly about unobservable mental content of others, such as beliefs, goals, and intentions. People often use this ability to understand the behavior of others as well as to predict future behavior. People even take this ability a step further, and use *higher-order theory of mind* by reasoning about the way others make use of theory of mind and in turn attribute mental states to different agents. One of the possible explanations for the emergence of the cognitively demanding ability of higher-order theory of mind suggests that it is needed to deal with mixed-motive situations. Such mixed-motive situations involve partially overlapping goals, so that both cooperation and competition play a role. In this paper, we consider a particular mixed-motive situation known as Colored Trails, in which computational agents negotiate using alternating offers with incomplete information about the preferences of their trading partner. In this setting, we determine to what extent higher-order theory of mind is beneficial to computational agents. Our results show limited effectiveness of first-order theory of mind, while second-order theory of mind turns out to benefit agents greatly by allowing them to reason about the way they can communicate their interests. Additionally, we let human participants negotiate with computational agents of different orders of theory of mind. These experiments show that people spontaneously make use of second-order theory of mind in negotiations when their trading partner is capable of second-order theory of mind as well.

Keywords Theory of mind · Cognitive hierarchy · Opponent modeling · Agent-based models · Simulation · Incomplete information · Negotiation

✉ Harmen de Weerd
h.a.de.weerd@rug.nl

[1] Institute of Artificial Intelligence, Faculty of Mathematics and Natural Sciences, University of Groningen, Groningen, The Netherlands

Published online: 10 December 2015

1 Introduction

In social settings, people often make predictions of the behavior of others by making use of their *theory of mind* [57]; they reason about unobservable mental content such as beliefs, desires, and intentions of others. Without this theory of mind, an individual is limited to reasoning only about behavior, such as in the sentence "Mary is looking in the drawer". Such individuals are said to have a *zero-order theory of mind*. *First-order theory of mind* allows agents to reason about unobservable mental content of others as well, and understand sentence like "Mary is looking in the drawer because she *believes* that there is chocolate in the drawer". People are also capable of taking this theory of mind ability a step further, and reason about way others are using theory of mind. Using *second-order theory of mind*, people understand sentences such as "Alice *believes* that Bob *knows* that Carol is throwing him a surprise party", and reason about the way Alice is reasoning about Bob's knowledge.

Behavioral experiments have demonstrated the human ability to make use of higher-order (i.e. at least second-order) theory of mind, both through tasks that require explicit reasoning about second-order belief attributions [1,3,48,56,74], as well as through strategic games [18,26,34,38,47,60,77]. According to the social brain hypothesis [21], the emergence of this higher-order theory of mind ability can be explained by an increased complexity of social life. However, different hypotheses point to different aspects of social life that would favor the emergence of higher-order theory of mind.

According to the Machiavellian intelligence hypothesis [7,73], the emergence of social cognition, which includes theory of mind, can be explained through a *competitive* advantage. According to this theory, higher-order theory of mind allows an individual to deceive and manipulate others more effectively. Our earlier research using agent-based models has confirmed that there are indeed competitive settings in which individuals benefit from the use of higher-order theory of mind [13]. However other agent-based models have shown that theory of mind is not always needed to deceive others. Rather, seemingly deceptive behavior may be a result of associative learning [19,39,54] or factors such as stress [67] rather than reasoning about the minds of others.

The Vygotskian intelligence hypothesis [49,72] suggests that the emergence of theory of mind can be explained through *social cooperation* rather than competition. The Vygotskian intelligence hypothesis would explain both the human capacity for theory of mind and the capacity to engage in altruistic cooperative action [6,29,65]. Our results from agent-based simulations in a communication game show that higher-order theory of mind can indeed help to reach a cooperative solution more quickly [14]. However, computational models have shown that many forms of cooperation can also emerge through simple mechanisms, without need for a cognitively demanding ability such as theory of mind [12,51,66].

Finally, a third hypothesis that specifically concerns higher-order theory of mind states that higher orders of theory of mind may be needed for *mixed-motive interactions* [71]. Such mixed-motive interactions involve both cooperative and competitive elements,[1] such as in negotiations [71]. Mixed-motive interactions can be understood as the task of sharing a pie [61]. Individuals cooperate to find ways to enlarge the pie they are sharing, while they also compete to obtain as large a share of the pie as possible for themselves. Theory of mind allows individuals to reason explicitly about the goals and beliefs of others. This ability may be crucial for an individual to balance cooperative and competitive goals in order

[1] Note that mixed-motive interactions are different from risk/benefit trade-offs such as the Stag Hunt. In risk/benefit trade-off games, players prefer the same *payoff-dominant* outcome (hunting stag in the Stag Hunt), while players in a mixed-motive situation have different preferences concerning the outcome of the game.

to successfully negotiate a larger pie to share, which includes a larger piece of pie for the individual himself.

In this paper, we investigate whether theory of mind allows agents to achieve better outcomes in mixed-motive interactions. Our earlier research into the effectiveness of higher-order theory of mind shows that in repeated one-shot interactions in the negotiation game Colored Trails, agents that made use of first-order and second-order theory of mind managed to negotiate a larger piece of pie for themselves than agents of a lower order of theory of mind, while no additional advantage for even higher orders of theory of mind was found [16].

In the current paper, we extend the agent model of [16] and investigate more realistic mixed-motive settings in which individuals engage in multiple rounds of negotiation. Using agent-based computational models, we simulate agents that alternate in making offers until an agreement is reached. We study mixed-motive situations through the influential Colored Trails setting, introduced by Grosz et al. [15,28,45,70],[2] which provides a useful test-bed to study how different aspects of mixed-motive settings change the interactions among agents. By comparing the performance of agents of different orders of theory of mind over a variety of different game boards, we then determine to what extent higher orders of theory of mind allow agents to make better offers. Next, we let human participants negotiate with these computational agents to show that participants indeed take advantage of the benefits of higher-order theory of mind reasoning, even when playing against computational agents.

The remainder of this paper is structured as follows. In Sect. 2, we provide an overview of literature related to the current work. In Sect. 3, we describe our version of the Colored Trails game in detail. Section 4 describes how this game is played by agents, and how theory of mind shapes the decisions of agents. The details concerning these theory of mind agents are presented in the form of a complete formal model in Sect. 5.

By simulating negotiations among computational agents, we determine to what extent higher orders of theory of mind provide agents with an advantage over trading partners without such abilities. The results of these simulation experiments, in which theory of mind agents of various orders of theory of mind negotiate among each other, are presented in Sect. 6. In Sect. 7, we add humans to the loop by letting human participants play against our computational theory of mind agents. In that section, we show that participants are capable of spontaneous use of second-order theory of mind when negotiating with a computational agent. Finally, Sect. 8 provides discussion and gives directions for future research.

2 Related work

In the literature, there are several approaches to bounded rationality and recursive modeling of the behavior of others that are related to the theory of mind agents we present in this study in different ways (see also Table 1). In behavioral economics, recursive modeling of the behavior of others can be modeled through iterated best-response models such as level-n theory [2,4,10,50,64], cognitive hierarchies [8], quantal response equilibria [46], and noisy introspection models [33]. In these models, an agent's level of sophistication is measured by the maximum number of steps of iterated reasoning the agent is capable of considering. Camerer et al. [8] find that over a range of non-repeated single-shot games such as the p-beauty contest and the traveler's dilemma, participants typically use few reasoning steps. On average, participants use an estimated 1.5 steps of iterated reasoning, which suggests that participants limit themselves to first-order theory of mind reasoning. In a meta-analysis of

[2] Also see http://coloredtrails.atlassian.net/wiki/display/coloredtrailshome/.

Table 1 Related research has focused mostly on single-shot interactions and on single games

Paper	Setting	Player interaction	Scenario variety
Devaine et al. [18]	Competitive	Single-shot	Single game
	Cooperative	Single-shot	Single game
Ficici and Pfeffer [24]	Mixed-motive	Single-shot	Randomized games
Franke and Galleazzi [27]	Randomized	Single-shot	Randomized games
Peled et al. [53]	Mixed-motive	3 rounds	2 games
Pynadath et al. [59]	Mixed-motive	Arbitrary length	Single game
Wright and Leyton-Brown [75]	Competitive	Single-shot	
Yoshida et al. [76]	Cooperative	10–20 actions	Single game

In contrast, the current work involves a mixed-motive setting in which we consider arbitrary length player interaction across an extensive number of randomized games

these types of games, Wright and Leyton-Brown [75] find evidence of participant behavior that is consistent with higher-order theory of mind reasoning. However, few players were found to be well-described as higher-level agents.

In the iterated reasoning models described above, a level-n agent assumes that all other agents are exactly one level of sophistication lower than himself, or that the distribution of lower level agents can be described with a fixed probability distribution. However, in repeated game settings, such assumptions can be detrimental to an agent [40]. The theory of mind agents we describe in the rest of this article are more similar to dynamic models of theory of mind, such as experience weighted attraction learning [9], recursive opponent modeling [30,32], interactive POMDPs [31], and game theory of mind [76]. In these approaches, agents adjust their level of recursive reasoning in reaction to the behavior of others. An agent of level k can consider others as being agents of any level up to and including level $k-1$. Such an agent does not observe the level of sophistication of others directly, but forms beliefs concerning the level of sophistication of others based on observed behavior.

These dynamic models of theory of mind reasoning show that over repeated trials, human participants can successfully increase their level of theory of mind reasoning. For example, Doshi et al. [20] use adjusted interactive POMDPs to model human behavior in repeated competitive single-shot games. They find that although humans generally reason at low levels of theory of mind, participants exhibit higher levels of reasoning in simpler settings. Yoshida et al. [76] evaluate the behavior of human participants in a sequential game variation on the cooperative Stag Hunt game. Using game theory of mind, they find evidence that participants make use of higher-order theory of mind reasoning.

Our work focuses on mixed-motive settings that involve both cooperative and competitive elements, such as in negotiations. As a result, our work is related to research into automated agents in negotiation applications [23,26,41,43–45,53,62]. In particular, several studies have previously investigated recursive reasoning in the Colored Trails setting under incomplete information about the preferences of other players.

For example, Ficici and Pfeffer [24] present a model for recursive reasoning in repeated single-shot negotiations in Colored Trails and use this model to determine to what extent human participants reason about other players in the game. They find that human participants engage in theory of mind reasoning, but that more complex models yield diminishing returns. Peled et al. [53] construct an agent model for revelation games, in which players can decide to truthfully reveal their goals before engaging in two rounds of negotiation. Peled et al. fit their

SIGAL agent model to participant data on two game boards and show that the SIGAL agent could outperform both human participants and equilibrium strategy models in negotiations with participants.

In the current work, we investigate a more open-ended type of bargaining in which agents negotiate until either an agreement is reached or one of the agents withdraws from negotiation. This setting allows agents to observe more of the behavior of their trading partner, which may reduce the benefit of reasoning about the mental content of others. Indeed, Pynadath et al. [59] show this effect in a simple negotiation setting. Pynadath et al. model theory of mind reasoning in a simple type of open-ended bargaining in *Sigma* (Σ), an integrated computational model of intelligent behavior that is grounded in a cognitive architecture [42], and find that the ability to make use of theory of mind is only marginally beneficial. In this paper, we investigate a much more complex negotiation setting in which no agent encounters the exact same negotiation game twice. In this setting, we show that the use of higher-order theory of mind can still be beneficial.

In addition, we take the role of learning into account for agents that are unable to use theory of mind reasoning. In previous work, the most basic agent typically assumes that the behavior of others is either fixed or can be modeled as noise. In contrast, our zero-order theory of mind agents make use of an associative learning technique [19,39], and continue to adjust their actions based on the behavior of others, but without any mental state modeling. By observing the behavior of higher-order theory of mind agents, a zero-order agent may therefore learn to behave as if he were a more sophisticated agent. Our agent model therefore explicitly takes the role of learning into account as an alternative to higher-order theory of mind reasoning.

The goal of our work is to identify settings in which there is an evolutionary incentive to reason using higher orders of theory of mind which could explain the emergence of human-like theory of mind abilities. However, although we model human-like theory of mind abilities, our goal is not to replicate actual human social behavior in the same way as agent-based simulation tools such as PsychSim [58] or to predict human theory of mind inferences like in Bayesian Theory of Mind [5]. Instead, we explicitly compare simple learning strategies that rely solely on modeling the behavior of others with more complex strategies that include theory of mind to determine the extent of their effectiveness. In this sense, our work also differs from formal methods such as (dynamic) epistemic logic [22,68] and epistemic game theory [35–37,52,55], which are used to study recursive reasoning about the knowledge of others from a prescriptive perspective. In contrast, our theory of mind agents typically construct an incorrect model of the beliefs of others. We determine to what extent reasoning at increasingly higher orders of theory of mind remains effective, even under these conditions.

The evolutionary advantage of higher-order theory of mind has recently received more attention. For example, Franke and Galeazzi [27] compare the evolutionary success of level-n agents in repeated randomly generated single-shot games. They find that level-n agents of increasingly higher levels continue to obtain an advantage over level-$(n-1)$ agents. Interestingly, their results also show that populations that only contain high-level reasoners can sometimes be invaded by level-0 agents. This means that under certain circumstances, populations that contain both low-level and high-level agents can be evolutionarily stable.

Recently, Devaine et al. [18] investigated the effectiveness of higher-order theory of mind using a model of meta-Bayesian agents that is closely related to our agent model. Using replicator dynamics, Devaine et al. determine whether Bayesian agents of a lower order of theory of mind can survive when faced with more sophisticated agents in both a competitive setting and a cooperative setting. In the competitive hide-and-seek setting, their Bayesian theory of mind agents benefit from the ability to make use of increasingly higher orders

of theory of mind. In this setting, only agents using the highest order of theory of mind survive in the population. In a cooperative setting, on the other hand, reasoning at higher orders of theory of mind is not always beneficial. Devaine et al. find that in the battle of the sexes setting, the population reaches an evolutionarily stable state when two-thirds of the population consist of second-order theory of mind agents while the remaining one-third of the population consists of first-order theory of mind agents.

In contrast to previous models that investigate the evolutionary advantage of higher-order theory of mind in relatively simple repeated single-shot games, we consider an open-ended negotiation setting in which agents negotiate an agreement over multiple rounds of offers. To ensure that our results are generalizable, each new negotiation is played on a game board that agents have never encountered before. By comparing the performance of theory of mind agents and agents that rely on simpler learning techniques, we aim to determine to what extent higher-order theory of mind reasoning allows agents to obtain better outcomes in mixed-motive situations such as negotiations. Through negotiations between human participants and computational agents, we also investigate to what extent participants take advantage of the benefits of higher-order theory of mind reasoning.

3 Colored Trails

To determine to what extent reasoning at higher orders of theory of mind results in better outcomes in mixed-motive situations, we compare performance of computational agents that negotiate in the setting of Colored Trails. Colored Trails is a board game designed as a research test-bed for investigating decision-making in groups of people and computer agents [28,45,70]. As a prototypical multi-issue bargaining situation, the Colored Trails setting can capture a wide variety of negotiation aspects. For example, many of the domains in the GENIUS framework [44] can be directly implemented as a Colored Trails setting. Our specific setting is similar to the one we used previously to test the effectiveness of higher-order theory of mind in single-shot negotiations [16]. The game is played by two players on a square board consisting of 25 tiles that are randomly assigned one of five possible colors, such as the board in Fig. 1. At the start of the game, each player receives a set of four colored chips, selected randomly from the same five possible colors as those on the board. Each player is initially located on the center tile of the board, indicted with the letter *S* in Fig. 1a. Players can move to a tile adjacent to their current location by handing in a chip of the same color as the destination tile. Figure 1b shows an example of one of the 5^{24} possible Colored Trails boards as well as a possible path across the board. A player following the path from location *A* to the white tile marked *B* would have to hand in one black chip, one gray chip, and one white chip.

Each player is also assigned a goal location, which is randomly drawn from the board tiles that are at least three steps away from the initial location (striped tiles in Fig. 1a). The goal of each player is to approach the goal as closely as possible. To reach that goal, players are allowed to trade chips among each other. This trading of chips in the Colored Trails setting represents a multi-issue bargaining situation, in which every color represents a different issue or task to overcome. Different paths from the initial location to the goal location on the board represent different ways of achieving the same goal, while each chip represents the means to complete a task or resolve an issue. In our specific Colored Trails setting, for example, each player always has at least three possible paths from the starting location to the goal location.

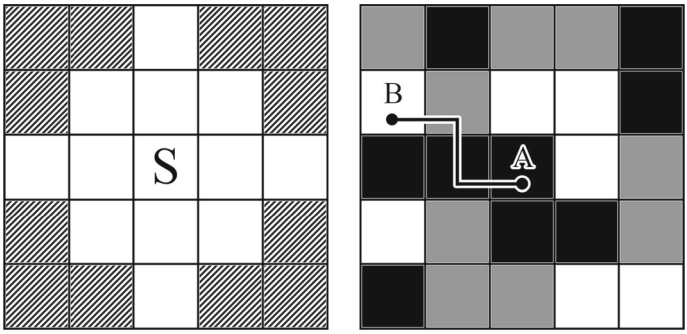

Fig. 1 The Colored Trails game is played on a 5 by 5 board. **a** Each player starts at the central tile *S* and receives a goal location drawn randomly from the striped tiles. **b** To follow the path from location *A* to location *B*, a player needs to hand in one *black*, one *gray*, and one *white* chip

In Colored Trails, players are scored based on their success in reaching their goal location. For each step a player takes towards his or her goal, the player receives 100 points. Any player that succeeds in reaching their goal receives an additional 500 points. Finally, any chip that has not been used to move around the board is worth an additional 50 points to its owner. This scoring ensures that players have the highest incentive to reach their goal location, but that they are also motivated to compete over control of unused chips.

Although players are scored based on how closely they approach their own goal, Colored Trails is not a strictly competitive game. Since a player may need a different set of chips to achieve his goal than his trading partner, there may be an opportunity for a cooperative trade, which allows both players to obtain a higher score. That is, although the score of a player is not influenced by how closely his trading partner reaches his or her goal location, players may still benefit from taking into account the goal of their trading partner. However, agents in our Colored Trails setup do not know the goal location of their trading partner from the start. In addition, each negotiation game is played on a new game board, which is randomly selected from one of the 5^{24} possible game boards, with new initial sets of chips and a new goal location. This means that players are very unlikely to see a given game setting more than once.

Trading among players takes the form of a sequence of alternating offers. The *initiator* makes an initial offer for a redistribution of chips. His trading partner then decides whether or not to accept this offer. If the offer is accepted, the proposed distribution of chips becomes final, the players move as close to their respective goal locations as possible, and the game ends. Alternatively, the trading partner may decide to withdraw from negotiations, which makes the initial distribution final. Finally, the trading partner may also decide to continue the game by rejecting the offer, and make his own offer for a redistribution of chips.

There are no restrictions on the offers that players can make. For example, a player is allowed to repeat an offer that has been previously rejected by his trading partner, or make an offer that he has previously rejected himself. In addition, our setting does not include a negotiation deadline. Instead, to prevent negotiations from taking an unbounded number of rounds to resolve, both players pay a 1 point penalty for each round of play. That is, when negotiations end after a total of five offers have been made, the final score of each player is reduced by five points. Note that this penalty is meant only to deter negotiations that last

indefinitely. As a result, this cost of negotiation has intentionally been kept low compared to the possible gains of negotiation.

In this paper, we investigate to what extent higher orders of theory of mind allow computational agents to make better offers. Based on our previous results in the one-shot variation of Colored Trails [16], we expect that theory of mind will provide agents with significant advantages over agents that are more limited in their theory of mind abilities. More specifically, we expect agents that are capable of a higher orders of theory of mind to be able to manipulate the beliefs of their trading partner in order to achieve higher individual scores than agents that are more limited in their theory of mind abilities. Additionally, we also expect that the presence of higher-order theory of mind agents in the negotiation has a positive effect on social welfare, as measured by the sum of the scores of the two negotiating agents. These expectations are captured by hypotheses H_1 and H_2.

Hypothesis H_1: First-order theory of mind agents obtain a higher score than zero-order theory of mind agents, while second-order theory of mind agents obtain a higher score than first-order theory of mind agents.
Hypothesis H_2: Social welfare, measured as the sum of the scores over the two agents, increases when the theory of mind abilities of either agent increases.

In Sect. 4, we describe the way computational agents play Colored Trails, and how the ability to reason about the goals of others influences the choices agents make. The formal description of these agents can be found in Sect. 5.

4 Theory of mind in Colored Trails

In this section, we describe the way theory of mind agents play Colored Trails. In our agent model, an agent achieves theory of mind by taking the perspective of his trading partner, and determining what his own decision would be if the agent had been in the position faced by his trading partner. Using the implicit assumption that his trading partner's thought process can be accurately modeled by his own thought process, the agent then predicts that his trading partner will make the same decision the agent would have made if the roles had been reversed.

In the remainder of this section, we describe how this process of perspective-taking results in different behavior for agents of different orders of theory of mind playing Colored Trails. The formal description of these theory of mind agents is presented in Sect. 5. We will use the shorthand ToM_k agent to indicate an agent that has the ability to use theory of mind up to and including the k-th order, but not beyond.

4.1 Zero-order theory of mind agent

The zero-order theory of mind (ToM_0) agent can model the behavior of his trading partner, but the ToM_0 agent is unable to attribute mental content to others. In particular, the ToM_0 agent is unable to represent that his trading partner wants to reach a certain goal location, and that the behavior of the trading partner is consistent with that desire. A ToM_0 agent is essentially fixated on his own piece of pie, and does not consider the piece of pie of other agents at all. Instead, the ToM_0 agent constructs zero-order beliefs about the likelihood that his trading partner will accept a certain offer. The ToM_0 agent bases these zero-order beliefs on observations of the behavior of the trading partner. For example, through repeated interaction, the ToM_0 agent will learn that offers that assign many chips to the trading partner

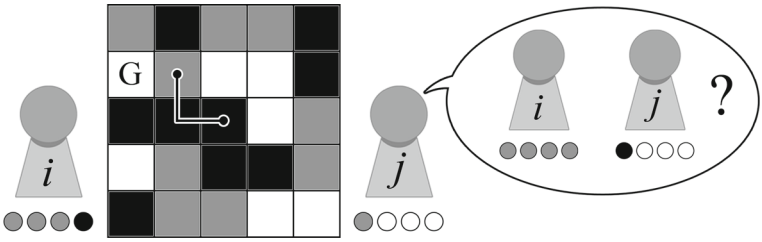

Fig. 2 Example of a negotiation setting in Colored Trails, in which agent j offers to trade the *black* chip owned by agent i against the *gray* chip owned by agent j. Agent i wants to move from the central square to his goal location G. With his initial set of chips, agent i can move two tiles towards his goal location, as shown by the *black* path

and few to the ToM_0 agent are more likely to be accepted, while offers that assign few chips to the trading partner and many to the ToM_0 agent himself are more likely to be rejected.

Using these zero-order beliefs, the ToM_0 agent can form an expectation about how his score will change if he were to make a particular offer, and select the offer that he assigns the highest expected value. This allows the ToM_0 agent to play the Colored Trails setting without attributing mental content to others. That is, although the zero-order beliefs of the ToM_0 agent will eventually reflect that his trading partner has a desire for owning chips, the ToM_0 agent does not explicitly represent such a desire.

The ToM_0 agent engages purely in *positional bargaining* [25], by only reasoning about specific offers that he believes his trading partner will accept, and that he is willing to accept himself. Because the ToM_0 agent has no theory of mind, he is unable to represent that his trading partner has interests that underlie the offers that his trading partner is willing to accept.

Example 0 Consider the example depicted in Fig. 2, and suppose agent i in this setting is a ToM_0 agent with goal location G. With his initial set of chips, agent i can take two steps towards his goal. This leaves agent i with two unused gray chips and one step away from his goal location. To reach his goal location, agent i needs one white chip, of which agent j has three.

A ToM_0 agent reasons only about the behavior of his trading partner and is unable to consider that his trading partner has goals or desires. In his very first game, a ToM_0 agent has no experience on which to base his prediction of his trading partner's behavior. Such an agent will offer the distribution of chips that would yield him the highest score, that is, he would ask to be given all chips. A ToM_0 agent quickly learns that asking for chips without offering anything in return is never successful.

By repeatedly playing the Colored Trails game across different game boards, a ToM_0 agent will come to learn that the more chips he offers his trading partner, the more likely it is that the offer will be accepted. For example, ToM_0 agent i could offer to exchange one of his gray chips for two white chips. This would allow agent i to reach goal location G with two spare chips. Alternatively, if ToM_0 agent i believes that the additional 100 points for two spare chips is not worth the risk of the offer being rejected, the ToM_0 agent may offer to exchange two gray chips for one white chip instead.

Figure 2 shows that agent j makes an offer to exchange his gray chip for the black chip owned by agent i. When a ToM_0 agent i receives this offer, he updates his beliefs concerning what offers agent j will accept in this particular game. Specifically, agent i lowers his beliefs that his trading partner will accept any offer that does not assign at least three white chips and one black chip to agent j. In subsequent rounds, ToM_0 agent i is therefore more likely

to make offers that assign his black chip to agent j than offers that assign gray chips to agent j.

4.2 First-order theory of mind agent

In addition to his zero-order beliefs, a first-order theory of mind (ToM_1) agent considers the possibility that his trading partner has beliefs and goals as well, which determine whether or not his trading partner will accept an offer. A ToM_1 agent therefore realizes that in order to get a large piece of pie for himself, it is essential to enlarge the pie as a whole. The ToM_1 agent is able to consider the game from the perspective of his trading partner, and decide what his action would be if he were in the position of that player.

Since each player wants to increase his own score through negotiation, each player reveals information about his goal location whenever he makes an offer. Although a ToM_1 agent does not know the goal location of his trading partner, the agent can learn the goal location of his trading partner through the offers he receives. In Example 1, we discuss this process in more detail.

Although the ToM_1 agent is able to consider his trading partner as a ToM_0 agent, the ToM_1 agent does not know the extent of the theory of mind abilities of his trading partner with certainty. Through repeated interactions, the ToM_1 agent may learn that his first-order beliefs fail to accurately model the behavior of his trading partner. If this happens, the ToM_1 agent may choose to play as if he were a ToM_0 agent.

Example 1 Consider the negotiation board shown in Fig. 2 and suppose agent i is a ToM_1 agent with goal location G. Using his first-order theory of mind, a ToM_1 agent can take the perspective of his trading partner to determine whether agent j would accept a given offer O. However, since agents only know their own goal location, agent i starts every game knowing only that his trading partner's goal location is three or four steps away from the center of the board.

To determine whether an offer O is likely to be accepted, ToM_1 agent i determines how the score of his trading partner would change by accepting offer O. Figure 3 shows this

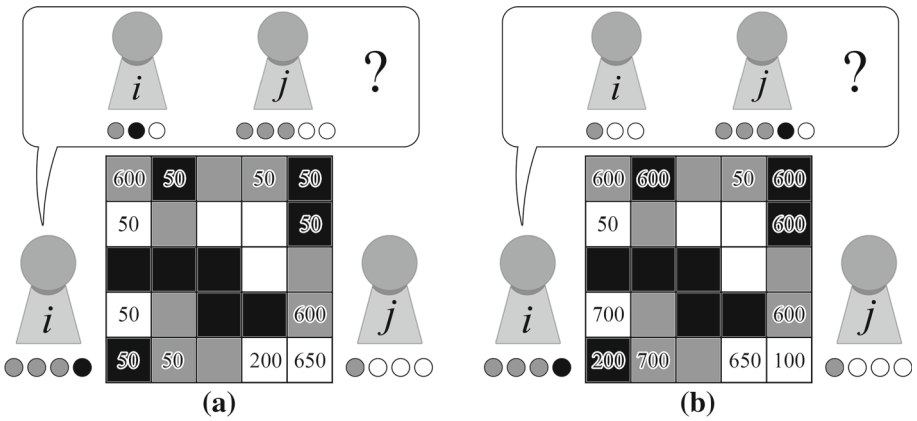

Fig. 3 If agent i is a ToM_1 agent, he reasons about the goal location of his trading partner. When agent i considers making an offer, he also considers how accepting that offer will change the score of agent j. *Panels* **a** and **b** show this change in score of agent j for two possible offers and for each possible goal location of agent j. In this case, agent i believes that his trading partner will only prefer offer (**a**) over offer (**b**) when his goal location is at the far *bottom right* of the board

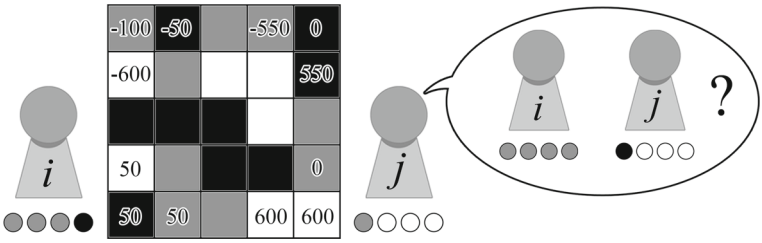

Fig. 4 Whenever a ToM_1 agent receives an offer from his trading partner, he updates his beliefs concerning his goal location. For each possible goal location, this figure shows how the score of agent j would change if agent i were to accept the offer. Since agent j would not make an offer that would decrease his own score, agent i concludes that the goal location of agent j is among those locations with a positive number

process for two possible offers. In Fig. 3a, agent i offers to exchange two of his gray chips against one white chip of agent j. In Fig. 3b, agent i offers to exchange two of his gray chips and his black chip for two white chips. Note that in both these situations, agent i can reach his goal location with no chips to spare. Although agent i has no preference for either one of these offers, agent i knows that the same may not be true for agent j. For each possible goal location of agent j, Fig. 3 shows how the score of agent j changes by accepting either one of these offers. For example, if the goal location of agent j is the bottom left square, offer Fig. 3a would increase his score by 50 points, while offer Fig. 3b would increase the score of agent j by 200 points. Without additional information about the goal location of his trading partner, agent i concludes that agent j is more likely to accept offer Fig. 3b than he is to accept offer Fig. 3a, since Fig. 3 shows that this would typically yield agent j a higher score.

By placing himself in the position of his trading partner, a ToM_1 agent can also interpret the offers he receives. Figure 4 shows an example in which agent j offers to trade the black chip owned by agent i against the gray chip owned by agent j. For each possible goal location, the figure shows how accepting this offer would change agent j's score. By placing himself in the position of agent j, agent i believes that agent j's goal location is one of the locations with a positive number. After all, for other locations, agent i would not have made the offer shown in Fig. 4. Furthermore, agent i considers it unlikely that the goal location of agent j is one of the locations that show a low number. For these locations, agent i reasons that he himself would have made a different offer.

4.3 Higher orders of theory of mind agent

Agents that are able to use orders of theory of mind beyond the first order consider the possibility that other agents take into account that others have beliefs and goals as well. A higher-order theory of mind agent reasons about the way his offers influence what his trading partner believes about his goal location. Such an agent may choose to select the offer that provides his trading partner with as much information as possible about his goal location. This way, a ToM_2 agent can inform his trading partner that he prefers a small piece of cherry pie over a large piece of chocolate pie, so that his trading partner can take this into consideration when making an offer. That is, higher orders of theory of mind allow agents to communicate their interest to their trading partner through the offers they make, and engage in *interest-based negotiation* [25].

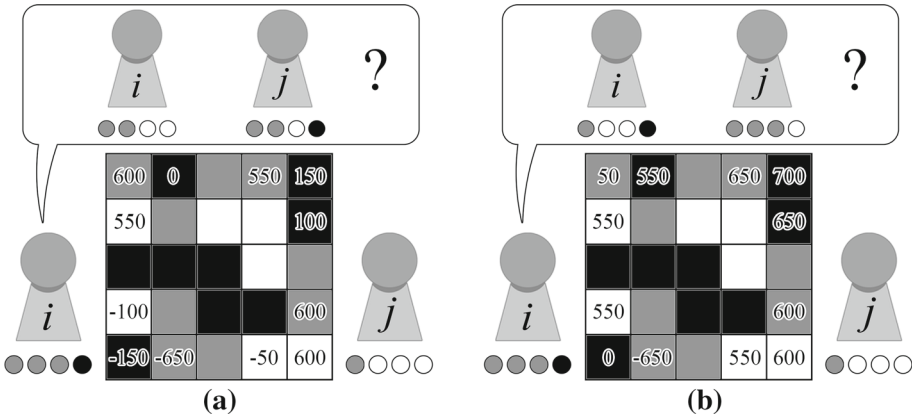

Fig. 5 A ToM_2 agent i takes into account that his trading partner does not know his goal location. *Panels* **a** and **b** show two possible offers that agent i could make, along with how this would the score of agent i for each possible goal location. Agent i can use this information to determine how his offers could change the beliefs of agent j concerning the goal of agent i

Higher orders of theory of mind also allow agents to manipulate the beliefs of trading partners of a lower order of theory of mind. For example, a higher-order theory of mind agent may construct an offer that gives his trading partner an incorrect impression of his goal location. Such an agent may exaggerate the value of the chips he already possesses and downplay the value of other chips in order to get a better deal. Whether a higher-order theory of mind agent decides to reveal his true goal location or attempt to manipulate the beliefs of his trading partner depends on what the agent believes to result in the highest score for himself.

As with the ToM_1 agent, a higher-order theory of mind agent does not know the extent of the theory of mind abilities of his trading partner. Instead, a ToM_k agent has $k+1$ hypotheses about the future behavior of his trading partner. While playing the Colored Trails game, the ToM_k agent continuously updates his beliefs concerning which of these hypotheses best fits the actual behavior of the trading partner. The details of this belief update procedure are described in Sect. 5.4.

Example 2 Consider the situation shown in Fig. 2 and suppose that agent i is a ToM_2 agent with goal location G. Using his second-order theory of mind, a ToM_2 agent believes that his trading partner may try to find our the goal location of agent i by interpreting the offers he makes.

While deciding what offer to make, a ToM_2 agent also takes into account how his offer influences his trading partner's beliefs concerning his goal location. Figure 5 shows two offers that ToM_2 agent i could make. For both these offers, the figure shows how accepting the offer would change the score of agent i for each possible goal location. Agent i knows that a ToM_1 trading partner can use this to obtain information about the goal location of agent i. For example, the offer shown in Fig. 5a excludes five locations as possible goal locations for agent i. This information would help a ToM_1 agent j to construct offers that are mutually beneficial. In contrast, the offer shown in Fig. 5b only excludes two locations, which would leave agent j with little information about the goal location of agent i.

5 Mathematical model of theory of mind

In the previous section, we presented the intuition behind agents negotiating in a Colored Trails setting using theory of mind. In this section, we discuss the implementation of computational agents that play according to this intuition. The agents described in this section is inspired by the theory of mind agents we used in [13] to investigate the effectiveness of theory of mind in competitive settings. This model is extended to allow for generalization over different stage games, and to allow for sequential games.

In our representation, a Colored Trails game is a tuple $\mathcal{CT} = \langle \mathcal{N}, \mathcal{D}, L, \pi_t^i, \pi_t^j, D_0 \rangle$, where:

- $\mathcal{N} = \{i, j\}$ is the set of agents;
- $\mathcal{D}$ is the set of possible distributions of chips;
- L is the set of possible goal locations;
- $\pi_t^i, \pi_t^j : L \times \mathcal{D} \rightarrow \mathbb{R}$ are the score functions for agents i and j respectively, such that $\pi_t^i(l, D)$ denotes the score of agent i at round t when his goal location is $l \in L$ and the chips are distributed according to distribution $D \in \mathcal{D}$; and
- $D_0 \in \mathcal{D}$ is the initial distribution of chips.

In addition, each agent i knows his own goal location l_i from the start of the game, while agent i does not know the goal location l_j of trading partner j. In the setting we describe, we therefore assume that each agent knows the set of possible offers $\mathcal{D}$, but has incomplete information about the preferences of his trading partner.

In our representation of the Colored Trails game, we focus on the negotiation aspect and ignore the task of finding routes between locations. This is reflected in the score functions π_t^i and π_t^j, which specify the maximum score agents i and j can achieve given some distribution of chips. This means that we assume that agents make no mistakes in finding routes between locations and that agents do not consider the possibility that their trading partner would make any mistake in finding these routes. Note that these assumptions do not imply that the agents have common knowledge about any aspect of the game. Rather, we assume that our theory of mind agents have no beliefs that contradict a common knowledge about such aspects of the game.

Over the course of the game, agents alternate in making offers, which results in a sequence of offers $\{O_0, O_1, \dots\}$. After the initial offer O_0 has been made, the agent who received the last offer O_t decides whether to accept the offer O_t, withdraw from negotiations, or make an offer O_{t+1} of his own. In the model description below, we will show formulas from the point of view of agent i. Formulas from the point of view of trading partner j are analogous.

In the following subsections, we describe in detail how theory of mind agents play the Colored Trails game. Sections 5.1, 5.2, and 5.3 describe the decision-making process of a ToM_0 agent, a ToM_1 agent, and a ToM_k agent ($k \geq 2$), respectively. Section 5.4 outlines how agents learn within the context of a single game, while Sect. 5.5 describes how agents learn across different games. For notational convenience, we will omit variables from functions if they can be derived from the context.

5.1 Model of zero-order theory of mind

The ToM_0 agent described in Sect. 4.1 does not form explicit beliefs about the mental content of others. Instead, the ToM_0 agent constructs zero-order beliefs $b^{(0)} : \mathcal{D} \rightarrow [0, 1]$ about the likelihood $b^{(0)}(O)$ that a certain offer O will be accepted by his trading partner. Using these zero-order beliefs, the ToM_0 agent can estimate the value of continuing negotiations by

making an offer O. That is, the expected value the ToM_0 agent assigns to making offer O is

$$EV_i^0(O, l_i, b^{(0)}) = b^{(0)}(O) \cdot \pi_t^i(l_i, O) + \left(1 - b^{(0)}(O)\right) \cdot \pi_t^i(l_i, D_0). \tag{1}$$

If the ToM_0 agent were to choose to continue negotiation, he would therefore randomly select an offer $O_t \in \mathcal{D}$ that he assigns the highest expected value. That is, he selects O_t^* such that

$$O_t^* := \arg\max_{O \in \mathcal{D}} EV_i^0(O, l_i, b^{(0)}). \tag{2}$$

However, making a counteroffer is not the only option available to the ToM_0 agent. After the initial offer of a game, an agent can also accept the previous offer O_{t-1} made by his trading partner. Finally, an agent may also decide to withdraw from negotiations, in which case the initial distribution becomes final.

The ToM_0 agent rationally decides which of the three options outlined above he will take. That is, the ToM_0 agent selects the option that he believes will yield him the highest score, as described in the ToM_0 response function:

$$ToM_{0i}(O_{t-1}, l_i, b^{(0)}) = \begin{cases} O_t^* & \text{if} \quad EV_i^{(0)}(O_t^*, l_i, b^{(0)}) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & \quad EV_i^{(0)}(O_t^*; l_i, b^{(0)}) > \pi_{t-1}^i(l_i, O_{t-1}) \\ \text{accept} & \text{if} \quad \pi_{t-1}^i(l_i, O_{t-1}) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & \quad \pi_{t-1}^i(l_i, O_{t-1}) \geq EV_i^{(0)}(O_t^*, l_i, b^{(0)}) \\ \text{withdraw} & \text{otherwise.} \end{cases} \tag{3}$$

Equation (3) shows that if the ToM_0 agent i believes that offer O_t^*, which he assigns the highest expected value, will yield him a higher score than either withdrawing or accepting the offer O_{t-1}, the agent will reject offer O_{t-1} and make counteroffer O_t^*. Alternatively, if the agent believes that offer O_t^* does not satisfy these conditions, but accepting the offer O_{t-1} would give him a higher score than withdrawing, the ToM_0 agent accepts the offer O_{t-1}. In all other cases, the ToM_0 agent withdraws from negotiation.

Note that although the ToM_0 agent observes the entire sequence of offers $\{O_0, O_1, \dots\}$, the agent does not explicitly use the entire sequence of offers to make a decision. Instead, the ToM_0 agent decides purely on the basis of his zero-order beliefs $b^{(0)}$, which describe his beliefs about the future behavior of the trading partner.

5.2 Model of first-order theory of mind

The use of first-order theory of mind allows a ToM_1 agent to put himself in the position of his trading partner to consider an offer from the perspective of his trading partner. To do so, the ToM_1 forms first-order beliefs $b^{(1)} : \mathcal{D} \rightarrow [0, 1]$ that represent what the zero-order beliefs of the ToM_1 agent would have been if he had been in the position of his trading partner. The ToM_1 agent can then attribute these beliefs to his trading partner to obtain a prediction of future behavior. That is, the ToM_1 agent considers the possibility that his trading partner believes that the probability of the ToM_1 agent accepting a given offer $O \in \mathcal{D}$ is $b^{(1)}(O)$.

A ToM_1 agent uses his first-order beliefs to predict his trading partner's behavior as using the ToM_0 response function described in Eq. (3). Given the goal location l_j of his trading partner, the ToM_1 agent calculates the expected value of making offer $O \in \mathcal{D}$ as

$$EV_i^{(1)}(l_j, O, l_i, b^{(1)}) = \begin{cases} \pi_t^i(l_i, D_0) & \text{if } ToM_{0j}(O, l_j, b^{(1)}) = \text{ withdraw,} \\ \pi_t^i(l_i, O) & \text{if } ToM_{0j}(O, l_j, b^{(1)}) = \text{ accept,} \\ \max\left\{\pi_{t+1}^i\left(l_i, \hat{O}_t^{(1)}\right), \pi_{t+1}^i\left(l_i, D_0\right)\right\} & \text{otherwise,} \end{cases} \quad (4)$$

where

$$\hat{O}_t^{(1)} = ToM_{0j}(O, l_j, b^{(1)}) \quad (5)$$

is the offer that the ToM_1 agent expects his trading partner to make in response to receiving offer O.

Equation (4) shows that the ToM_1 agent looks further ahead into the negotiation process than the ToM_0 agent. The ToM_0 agent only forms beliefs about whether or not his trading partner will accept an offer, the ToM_1 agent can also make a prediction about what counteroffer his trading partner could make, and whether the ToM_1 agent himself would accept this counteroffer. Since the game is sequential, this results in the ToM_1 agent looking one step further ahead into the negotiation process.

The sequential nature of the game also means that a player may change his beliefs after receiving an offer O_{t-1}, but before deciding whether or not to make a counteroffer O_t. In our agent model, the ToM_1 agent takes this belief update, which will be described in detail in Sect. 5.4, into account. When deciding on whether to make offer $O \in \mathcal{D}$, the ToM_1 agent determines how making this offer O would change his zero-order beliefs if he had been in the position of his trading partner, and makes further calculations using the adjusted first-order beliefs $U(b^{(1)}, O)$ (see Eq. (12)).

Since agents do not know the goal location l_j of their trading partner from the start, a ToM_1 agent cannot calculate Eq. (4) directly. Instead, the ToM_1 agent forms beliefs about the goal location of his trading partner in the form of a probability distribution $p^{(1)} : L \to [0, 1]$, so that the ToM_1 agent believes that the likelihood of his trading partner having goal location $l \in L$ is $p^{(1)}(l)$.

Although the ToM_1 agent is capable of using theory of mind, the ToM_1 agent considers the possibility that his first-order beliefs $b^{(1)}$ may not accurately predict the behavior of his trading partner. In this case, the ToM_1 agent may decide to rely on his zero-order beliefs $b^{(0)}$ instead. To model the ToM_1 agent's uncertainty concerning the appropriateness of the use of first-order theory of mind, the ToM_1 agent has a confidence variable $c_1 \in [0, 1]$, which indicates how much confidence the ToM_1 agent places in the predictions of first-order theory of mind. When deciding on the expected value of making an offer O, the ToM_1 agents weighs the predictions of first-order and zero-order theory of mind accordingly. In summary, a ToM_1 agent i calculates the expected value of making a given offer $O \in \mathcal{D}$ through

$$EV_i^{(1)}(O, l_i, b^{(0)}, b^{(1)}, p^{(1)}, c_1) = (1 - c_1) \cdot EV_i^{(0)}\left(O, l_i, b^{(0)}\right) + c_1 \cdot \sum_{l \in L} p^{(1)}(l) \cdot EV_i^{(1)}\left(l, O, l_i, U(b^{(1)}, O)\right). \quad (6)$$

Given these values, the ToM_1 agent randomly selects an offer $O_t^* \in \mathcal{D}$ that maximizes his expected value as a suitable counteroffer. Similar to the way the ToM_0 agent decides, the ToM_1 agent decides to accept, withdraw, or make a counteroffer based on what he expects will yield him the highest score.

$$ToM_{1i}\left(O_{t-1}, l_i, b^{(0)}, p^{(1)}, b^{(1)}\right) = \begin{cases} O_t^* & \text{if } EV_i^{(1)}(O_t^*) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & EV_i^{(1)}(O_t^*) > \pi_{t-1}^i(l_i, O_{t-1}) \\ \text{accept} & \text{if } \pi_{t-1}^i(l_i, O_{t-1}) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & \pi_{t-1}^i(l_i, O_{t-1}) \geq EV_i^{(1)}(O_t^*) \\ \text{withdraw} & \text{otherwise.} \end{cases} \quad (7)$$

First-order theory of mind benefits the ToM_1 agent in two ways. Firstly, theory of mind allows the agent to gain information about the goal location of his trading partner through the offers he receives. He does so by determining how consistent a possible goal location is with the offer his trading partner has made (see Sect. 5.4). Secondly, the ToM_1 agent takes into account that making an offer O changes the beliefs and behavior of his trading partner. This may allow the ToM_1 agent to make an offer O_t that he expects his trading partner to reject, with the intention of causing his trading partner to make an offer O_{t+1} that the ToM_1 agent is willing to accept.

5.3 Model of higher-order theory of mind

Agents that are able to use orders of theory of mind beyond the first can use this ability to attempt to manipulate the beliefs of lower orders of theory of mind to obtain an advantage. For example, a second-order theory of mind agent can use his understanding of first-order theory of mind agents to create an offer that signals his goal location to the trading partner as clearly as possible. Alternatively, the ToM_2 agent can adjust his offer to give his trading partner false information about his goal location.

Each additional order of theory of mind allows an agent to consider an additional model of opponent behavior. These models are constructed analogously to first-order theory of mind, and include additional beliefs $b^{(k)}$, location beliefs $p^{(k)}$, and a confidence c_k in kth-order theory of mind. Based on the application of kth-order theory of mind, the ToM_k agent formulates the expected value of making an offer $O \in \mathcal{D}$, given that his trading partner has goal location l, as

$$EV_i^{(k)}(l, O) = \begin{cases} \pi_{t-1}^i(l_i, D_0) & \text{if } ToM_{(k-1)\,j}(O) = \text{ withdraw,} \\ \pi_t^i(l_i, O) & \text{if } ToM_{(k-1)\,j}(O) = \text{ accept,} \\ \max\left\{\pi_{t+1}^i\left(l_i, \hat{O}_t^{(2)}\right), \pi_{t+1}^i(l_i, D_0)\right\} & \text{otherwise,} \end{cases} \quad (8)$$

where

$$\hat{O}_t^{(2)} = ToM_{1\,j}(O, l_j, b^{(1)}) \quad (9)$$

is the offer that the ToM_2 agent expects his trading partner to make in response to receiving offer O.

These expected values are then combined with the expected values of lower orders of theory of mind according to

$$EV_i^{(k)}(O) = (1 - c_k) \cdot EV_i^{(k-1)}(O) + c_k \cdot \sum_{l \in L} p^{(k)}(l) \cdot EV_i^{(k)}(l, O)). \quad (10)$$

This yields the kth-order theory of mind response function

$$ToM_{ki}(O_{t-1}) = \begin{cases} O_t^* & \text{if } EV_i^{(k)}(O_t^*) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & EV_i^{(k)}(O_t^*) > \pi_{t-1}^i(l_i, O_{t-1}) \\ \text{accept} & \text{if } \pi_{t-1}^i(l_i, O_{t-1}) > \pi_{t-1}^i(l_i, D_0) \text{ and} \\ & \pi_{t-1}^i(l_i, O_{t-1}) \geq EV_i^{(k)}(O_t^*) \\ \text{withdraw} & \text{otherwise.} \end{cases} \tag{11}$$

5.4 Learning from observations

Agents form beliefs about the likelihood that their trading partner will accept a given offer. During negotiation, agents update these beliefs based on the offers their trading partner makes. Note that whether or not an agent will accept an offer depends on the specific game board, the distribution of chips, the agent's goal location, and the history of offers made during the negotiation process. However, since our Colored Trails setting spans 5^{24} possible game boards in addition to possible distributions of initial sets of chips and goal locations, it is not feasible for a ToM_0 agent to form meaningful beliefs that are specific for each possible game setting, let alone for each possible history of offers. In this section, we therefore present a way in which ToM_0 agents can still generalize the behavior of their trading partner using a simple learning heuristic.

An agent's zero-order belief $b^{(0)}$ specifies that the agent believes that the probability that his trading partner will accept a given offer $O \in \mathcal{D}$ is $b^{(0)}(O)$. Whenever he receives an offer O_{t-1} from his trading partner, the ToM_0 agent updates his beliefs to reflect that he considers it less likely that his trading partner would accept certain offers. More precisely, the ToM_0 agent decreases his belief that his trading partner will accept an offer O when offer O assigns more chips of some color c to the agent himself than offer O_{t-1} does. For example, suppose that the trading partner makes an offer O_{t-1} that assigns 4 blue chips to agent i. Agent i then decreases his belief that the trading partner will accept any offer that assigns 5 or more blue chips to agent i himself.

The belief update as a result of receiving an offer O_{t-1} from the trading partner is represented by $U(b^{(0)}, O_{t-1})$, which is defined as

$$U(b^{(0)}, O_{t-1})(O) = (1 - \lambda)^m \cdot b^{(0)}(O), \tag{12}$$

where m is the number of colors for which the offer O assigns fewer chips to the trading partner than the offer O_{t-1}, and $\lambda \in [0, 1]$ is an agent-specific learning speed.

A similar update takes place when the trading partner rejects an offer made by agent i. When the offer O_t made by agent i is rejected, the agent updates his beliefs to reflect that he believes it to be less likely that his trading partner will accept an offer O that assigns at least as many chips of a given color c to the agent as offer O_t does. The belief update as a result of receiving an offer O_{t-1} from the trading partner is represented by $U(b^{(0)}, O_{t-1})$, which is defined as

$$U^R(b^{(0)}, O_t)(O) = (1 - \lambda)^{m'} \cdot b^{(0)}(O), \tag{13}$$

where m' is the number of colors for which the offer O assigns at least as many chips to agent i as the offer O_t, and $\lambda \in [0, 1]$ is an agent-specific learning speed.

Agents that make use of theory of mind also update their beliefs concerning the goal location of their trading partner in response to receiving an offer O_{t-1} from their trading partner. By putting himself in the position of his trading partner, a ToM_k agent believes it to be impossible that his trading partner has a goal location $l \in L$ for which $\pi_{t-1}^j(l, O_{t-1}) \leq$

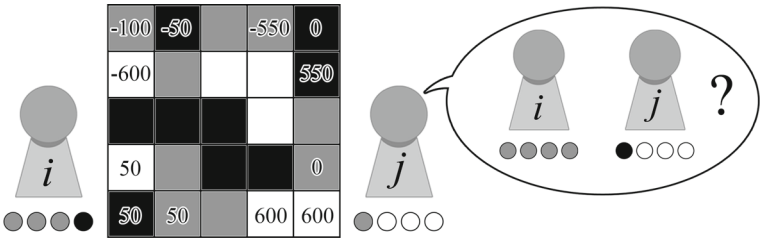

Fig. 6 Example of a Colored Trails game in which agent j offers to trade his *gray* chip against the *black* chip of agent i. Using first-order theory of mind, agent i calculates in what way accepting this offer will affect the score of agent j, for each possible goal location. The higher the increase in score, the more likely agent i considers the location to be the goal location of agent j

$\pi^j_{t-2}(l, D_0)$. For otherwise, this would mean that the trading partner has made an offer that would yield him a lower score than withdrawing from negotiation. For other locations $l \in L$, the agent adjusts his beliefs based on the expected increase in the score of the trading partner if the offer O_{t-1} would be accepted. That is, after observing the offer O_{t-1} from his trading partner, the ToM_k agent updates his location beliefs $p^{(k)}$ so that

$$p^{(k)}(l) := \begin{cases} 0 & \text{if } \pi^j_{t-1}(l, O_{t-1}) \leq \pi^j_{t-2}(l, D_0) \\ \beta \cdot p^{(k)}(l) \cdot \frac{1+EV_i^{(k-1)}(O_{t-1})}{1+\max_{O \in \mathcal{D}} EV_i^{(k-1)}(O)} & \text{otherwise,} \end{cases} \tag{14}$$

where β is a normalizing constant. This update increases the beliefs assigned to locations for which the offer O_{t-1} made by the trading partner receives a high expected value. These are offers that are closer to the offer that the ToM_k agent would have made himself if he had been a ToM_{k-1} agent in the position of his trading partner.

Example 3 Figure 6 shows an example of the process of updating location beliefs for a ToM_1 agent. In this example, agent j offers to trade the black chip owned by agent i against the gray chip owned by agent j. Agent i interprets this offer by calculating the change in score for agent j if agent i were to accept the offer, for each possible goal location of agent j. In Fig. 6, these changes in scores are shown on the corresponding locations. For example, if the goal location of agent j is the tile in the bottom right corner, the score of agent j would increase by 600 points if agent i were to accept the offer.

Since making an offer decreases the score of each agent by 1 point, agent i only makes offers that would increase his own score. By putting himself in the position of his trading partner, agent i therefore believes that agent j also only makes offers that increase the score of agent j. Agent i considers it impossible for any location with zero or negative score to be the goal location of his trading partner. That is, when agent i receives offer O, for each possible location $l \in L$ with $\pi^j_{t-1}(l, O) \leq \pi^j_{t-2}(l, D_0)$, agent i sets $p^{(1)}(l) = 0$.

For the remaining locations l, agent i determines what offer O' he would have made himself, and compares how this relates to the offer O that his trading partner actually made. For example, if the goal location of agent j is the bottom left tile, accepting the offer of agent j would only increase his score by 50. However, for this goal location, agent j could have made a better offer. If agent j would have offered to exchange a white chip for the black chip of agent i, he could have increased the score of agent j by 150. As a result, agent i believes that it is unlikely that the goal location of agent j is the bottom left tile. On the other hand, agent i considers it very likely that the goal location of his trading partner is one of

the locations with a number higher than 500, such as the bottom right tile. If agent i were to accept the offer made by agent j, agent j would be able to reach any of these locations.

At the same time as updating his location beliefs, the ToM_k agent also updates his confidence in kth-order theory of mind c_k to reflect how well the agent feels the kth-order theory of mind model fits the behavior of his trading partner. This is achieved through

$$c_k := (1-\lambda)\cdot c_k + \lambda \cdot \sum_{l\in L} p^{(k)}(l)\cdot \frac{1+EV_i^{(k)}(O_{t-1})}{1+\max_{O\in\mathcal{D}} EV_i^{(k)}(O)}. \tag{15}$$

Using this update, the agent therefore assigns a higher confidence to orders of theory of mind that assign a high expected value to the offer O_{t-1} made by the agent's trading partner compared to the offer that the agent would have selected himself is he had been a ToM_{k-1} agent in the position of his trading partner. Unlike the way location beliefs are updated, confidences are updated using adaptive expectations. This is because agents may change the order of theory of mind at which they reason over the course of a negotiation, while they are unable to change their goal location.

Many of the belief updates described in this section make use of learning speed parameter λ. The agent's learning speed is a fixed parameter that represents the degree to which the agent adjusts his beliefs in response to behavior of his trading partner. In addition to the order of theory of mind at which an agent is reasoning, an agent's learning speed λ also influences his negotiation strategy. For example, a ToM_0 agent with a high learning speed believes that his trading partner is unwilling to accept any offers other than the one the trading partner makes himself. Such a ToM_0 agent is less likely to make a counteroffer and more likely to withdraw from negotiations or accept the offer of his trading partner. On the other hand, a ToM_0 agent with learning speed $\lambda = 0$ does not adjust his behavior at all. Instead, such an agent will keep making the same offer until a successful trade is made.

Following [13], theory of mind agents do not attempt to model the learning speed λ of other agents. Instead, an agent makes use of his own learning speed when updating the beliefs he assigns to his trading partner. As a result, the beliefs that a theory of mind agent attributes to his trading partner are generally incorrect, unless both agents have the same learning speed.

5.5 Learning across games

Theory of mind allows agents to view the game from the perspective of their trading partner. This provides theory of mind agents with a way to generalize the behavior of the trading partner across the 5^{24} possible game boards (ignoring initial sets of chips). However, ToM_0 agents do not have the ability to reason about the goals of the other. Instead, ToM_0 agents reason only about the behavior of their trading partner. In this section, we discuss how a ToM_0 agent can generalize the behavior of the trading partner across different games without the use of theory of mind. Note that learning discussed in this section determines how the ToM_0 agent constructs his zero-order beliefs at the start of a negotiation. Over the course of a negotiation, agents perform additional belief updates as described in Sect. 5.4.

In our setting, the ToM_0 agent generalizes across games by classifying offers by the number of chips that are transferred from the agent to his trading partner, and the number of chips that are transferred from the trading partner to the agent himself. This allows agents to distinguish, for example, between an offer that trades one red chip for one blue chip and an offer that trades two red chips for two blue chips. However, across different games, the agent does not distinguish between an offer that trades one red chip for one blue chip and an offer that trades one green chip for one yellow chip. Since agents in our setting own an

initial set of four chips, this generalization causes the ToM_0 agent to distinguish 25 classes of offers. Nevertheless, a separate pilot study indicated that this simple heuristic allowed agents to make mutually beneficial offers after a short learning period.

At the start of each game, the agent's zero-order beliefs $b^{(0)}(O)$ about the probability that the trading partner will accept a given offer $O \in \mathcal{D}$ is set to the observed frequency with which offers that transfer the same number of chips from the agent to the trading partner and the same number of chips from the trading partner to the agent have been accepted by the trading partner in the past. For example, if a ToM_0 agent has made 250 offers in which two chips were transferred to the trading partner and one chip to the agent, of which 220 have been accepted by the trading partner, the ToM_0 agent assigns a probability of 88% that his trading partner will accept an offer to trade two green chips owned by the agent against one red chip owned by the trading partner at the start of the game. Over the course of the negotiation process, this belief can still change, as described in Sect. 5.4.

6 Simulation results

We performed simulations where the theory of mind agents described in Sect. 5 played the Colored Trails setting described in Sect. 3. Pairs of agents played repeated negotiation games, where each individual game is played on a different board with different sets of initial chips and different goal locations. Through the simulations, agents of different orders of theory of mind are confronted with trading partners with varying preferences and negotiation strategies. In each new game, agents started by reasoning at the highest order of theory of mind available to them. For example, a ToM_2 agent always started the game by taking into account the beliefs his trading partner might have about his own beliefs.

To ensure that all agents had an incentive to negotiate to increase their score, games in which some agent could reach his goal location with the initial set of chips without trading were excluded from analysis. Additionally, the first 200 games were considered to be a setup phase for the zero-order beliefs of agents, which were initialized at 1. That is, at the start of a simulation, an agent believes that any offer will be accepted. During the first 200 games, agents may learn, for example, that an "offer" that consists of requesting the trading partner's full set of chips is unlikely to be accepted. Similarly, higher-order theory of mind agents learn that their trading partner knows that such offers are unlikely to be accepted. The results from these 200 training games were not included in analysis.[3]

The figures in this section show the average change in score as a result of negotiation, which is calculated as the average difference between an agent's final score after negotiation ended and his initial score at the start of negotiation. Although agents never accept an offer that decrease their score, negative scores are possible when agents take many rounds. In these cases, the cost of negotiation can outweigh the benefits of a mutually beneficial trade. Negotiation scores were averaged over 10 runs of 1,000 consecutive Colored Trails games, each on a different game board. Although negotiations could take infinitely long in theory, games that continued for more than 100 rounds of offers were considered to be unsuccessful. In this case, the initial situation became final, and both agents incurred the cost of 100 rounds of play. In our model, agents were unable to reason about this limit, and negotiated as if this limit did not exist. With one exception (see Sect. 6.1), the average length of a negotiation was at most 15 rounds. The limit of 100 rounds of play therefore did not influence the length of the negotiation in general.

[3] Increasing the length of the training phase did not alter our results.

Although the agents alternate in making offers, so that both agents make offers to their trading partner, previous research into negotiation suggests that the opening bid of a negotiation can serve as an anchor for the entire negotiation process, making the first bid of a game especially influential [61,63,69]. Because of this, we differentiate between results for *initiators*, who make the first offer in every game, and *responders*.

In the following subsections, we separate results for the competitive and cooperative aspects of negotiation in Colored Trails. In Sect. 6.1, we present the individual performance of agents, which shows how well agents compete. Section 6.2 focuses on the cooperative element of negotiation, and describes the effect of theory of mind on the combined score of the agents in the Colored Trails setting. We conclude each of these sections with a short summary of the results.

6.1 Individual performance results

In this section, we describe the individual performance of theory of mind agents when negotiating in Colored Trails. By comparing how large a piece of pie the agents involved in Colored Trails end up with, we can determine how theory of mind influences the competitive abilities of agents. Throughout this section, we present graphs that show the scores of agents of various orders of theory of mind. The standard error of these agent scores was never higher than 9.12. As a result, a difference in score of at least 24 points is significant at $\alpha = 0.01$.

Figure 7 shows the average negotiation scores of a ToM_0 initiator and a ToM_0 responder negotiating with each other, as a function of the learning speeds of the two agents. In these figures, a lighter color indicates a higher score. As a visual aid, the plane of zero performance appears as a semi-transparent plane in these figures. Figure 7 shows that even though ToM_0 agents are unable to reason explicitly about the goals and desires of their trading partner, they are often able to increase their score through negotiation. The ToM_0 agents only fail to reach a positive negotiation score when both agents have learning speed $\lambda = 0$. An agent with zero learning speed does not adjust his behavior in response to his trading partner, but repeats the same offer until his trading partner makes an acceptable offer. That is, an agent

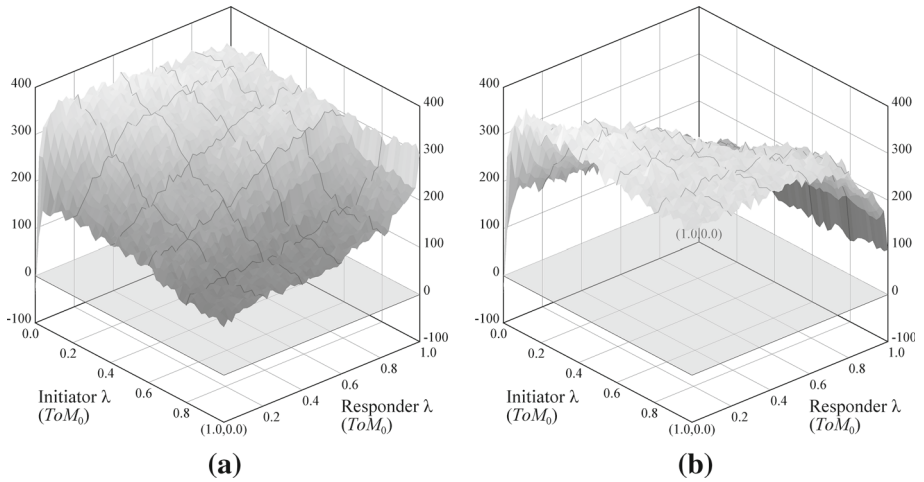

Fig. 7 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

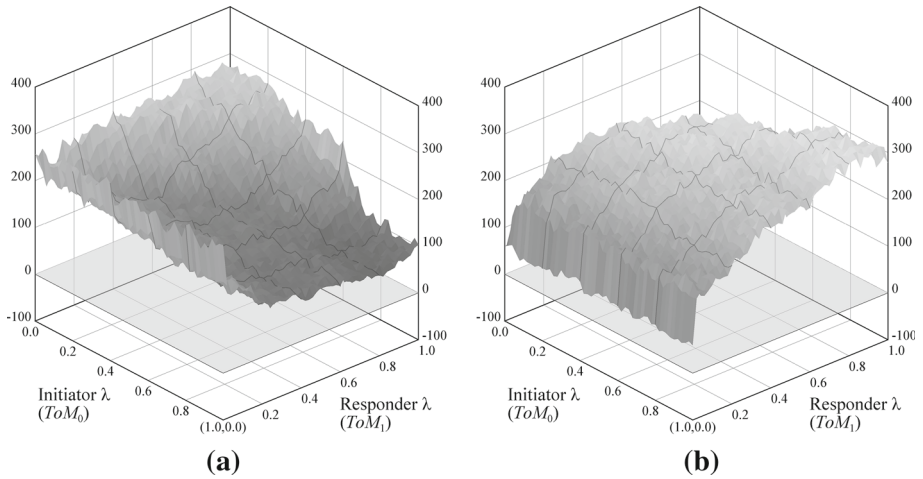

Fig. 8 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

with zero learning speed expects his trading partner to adjust to his position, while the agent is unwilling to offer any alternatives himself. When both ToM_0 agents follow this strategy and neither is willing to accept the initial offer of their trading partner, they will be unable to reach an agreement and only carry the burden of a failed negotiation.

Despite the fact that ToM_0 agents with a lower learning speed tend to engage in negotiations that take more turns, the results in Fig. 7 also show that for ToM_0 agents, negotiation score increases as their own learning speed decreases, unless the trading partner has learning speed $\lambda = 0$. This means that there is an evolutionary pressure on ToM_0 agents to decrease their learning speed, and adjust the offers they make as slowly as possible. However, this eventually results in the worst possible outcome in which negotiation fails.

Negotiation failure does not occur when a ToM_0 initiator negotiates with a ToM_1 responder. Figure 8 shows that for every combination of learning speeds of the ToM_0 initiator and the ToM_1 responder, both agents receive a positive score on average. However, the evolutionary pressure to reduce learning speed still exists for the ToM_0 initiator. A ToM_0 initiator receives the largest piece of pie when his learning speed is $\lambda = 0$, in which case he leaves only a small piece of pie for the ToM_1 responder. This means that although the presence of the ToM_1 responder prevents negotiation failure, the ToM_0 initiator benefits most.

Figure 9 shows a similar pattern when the roles are reversed, so that a ToM_1 initiator and a ToM_0 responder play Colored Trails. The ToM_0 responder performs best when his learning speed is zero, while the ToM_1 initiator prefers a higher learning speed. This allows the agents to negotiate successfully, with the ToM_0 responder receiving the most benefit. The presence of a ToM_1 agent yields a larger pie for the negotiating agents to share, but it is the ToM_0 agent that receives the largest piece.

Figure 9 also shows that when the ToM_1 initiator has a learning speed $\lambda \leq 0.4$, the score of both agents is zero. In these cases, the ToM_1 initiator withdraws from negotiation instead of making an initial offer. The reason for this is that the ToM_1 agent attributes his own learning speed to his trading partner. A ToM_1 agent with zero learning speed predicts that his trading partner will keep repeating the same offer until the ToM_1 agent makes an acceptable offer. This causes the ToM_1 agent to believe that the likelihood of finding a mutually beneficial

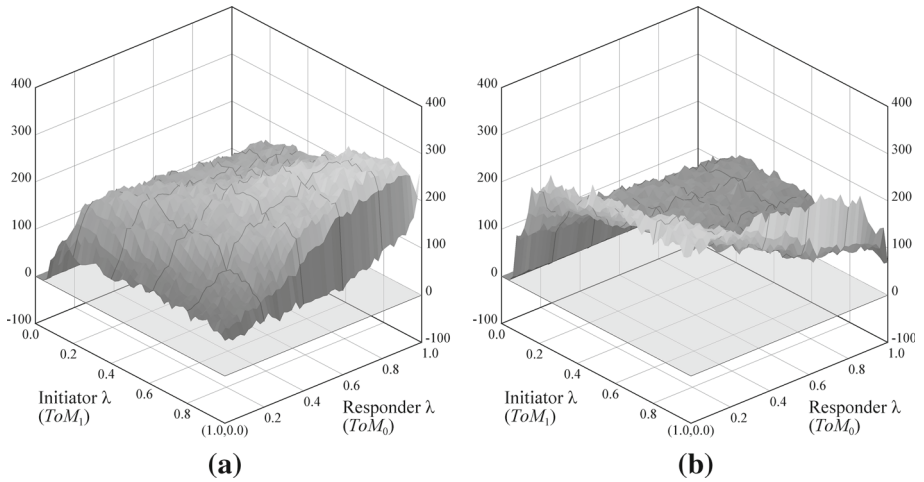

Fig. 9 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

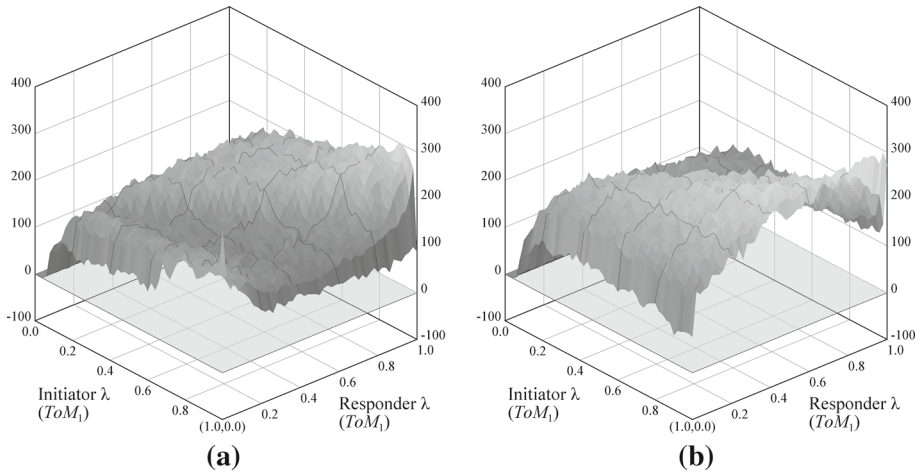

Fig. 10 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

trade is not worth the cost of negotiating. As a result, a ToM_1 initiator with learning speed $\lambda = 0$ withdraws from negotiation before making the initial offer.

Figure 10 shows the negotiation scores of a ToM_1 initiator and a ToM_1 responder negotiating in Colored Trails. The figures show symmetry around the line of equal learning speeds that indicates that the ToM_1 agent with the lower learning speed generally receives the largest piece of the pie. A ToM_1 agent with a higher learning speed attributes this learning speed to his trading partner and expects that his offers will influence the behavior of his trading partner more strongly. This also leads a ToM_1 agent to believe that his trading partner is quick to conclude that a negotiation will be unsuccessful. To prevent his trading partner from withdrawing from negotiations, the ToM_1 agent makes offers that he believes to be more beneficial for his trading partner at the expense of his own score. This puts evolutionary

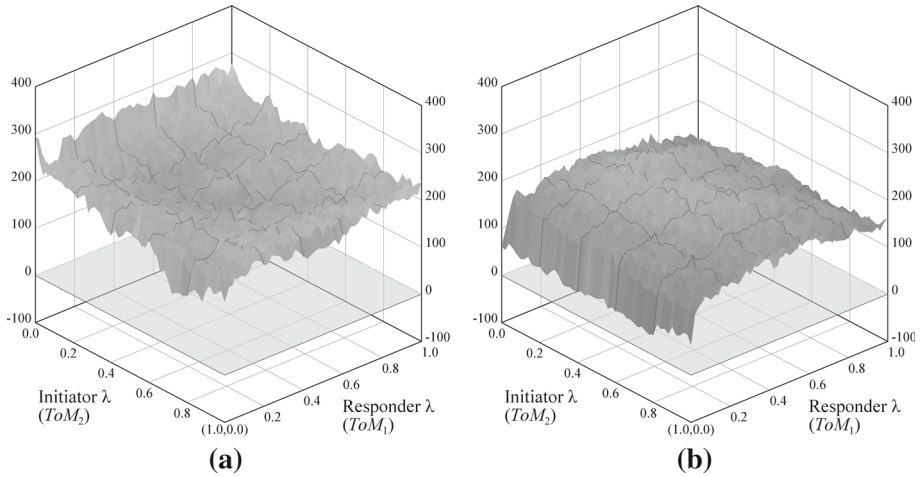

Fig. 11 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

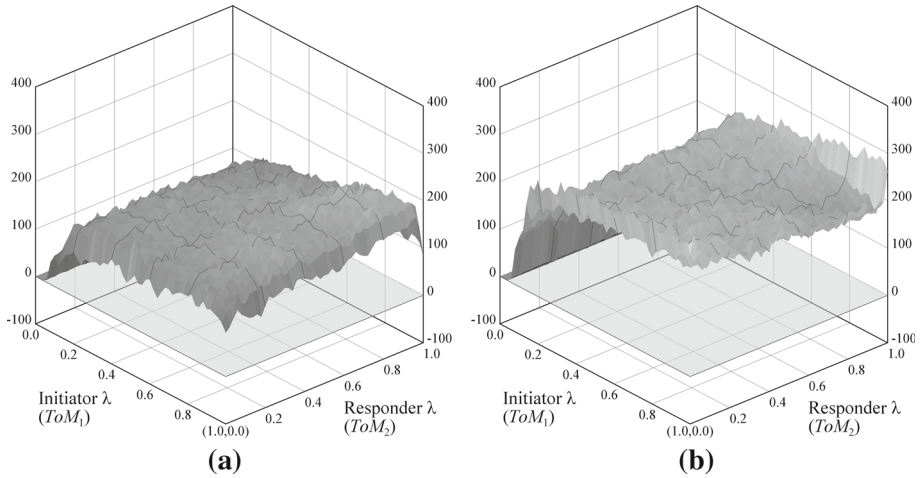

Fig. 12 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

pressure on ToM_1 agents to lower their learning speed. However, since ToM_1 agents perform poorly when their learning speed falls below $\lambda = 0.2$, the evolutionary pressure on ToM_1 agents is to have learning speeds close to $\lambda = 0.2$.

When a ToM_2 agent negotiates with a ToM_0 agent, the results are visually indistinguishable from the situation in which a ToM_1 agent negotiates with a ToM_0 agent. That is, when playing against a ToM_0 agent, the ability to make use of second-order theory of mind does not provide an agent with benefits additional to the use of first-order theory of mind. Note, however, that the ability to make use of second-order theory of mind is likely to require additional resources.

Figures 11, 12 show the performance of ToM_1 agents and ToM_2 agents that negotiate with each other. The graphs show that the ToM_2 agent typically has a higher score than his ToM_1

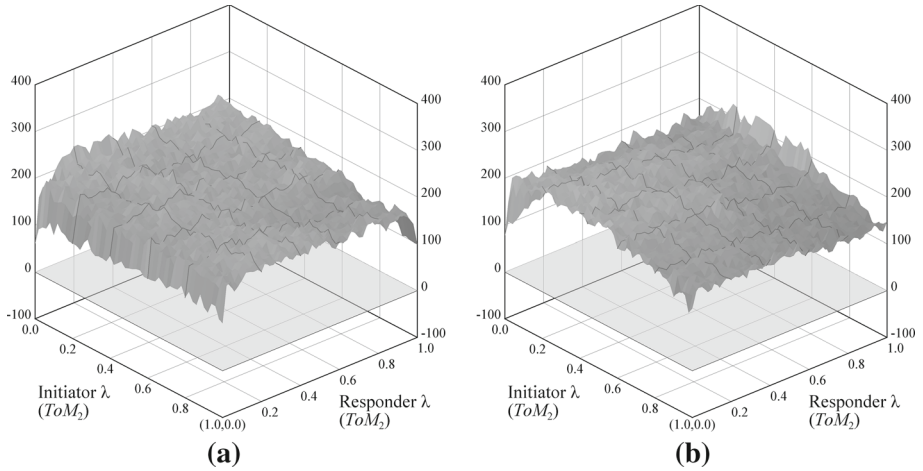

Fig. 13 Average negotiation score of a ToM_i initiator (a) and a ToM_j responder (b) negotiating with each other as a function of their respective learning speeds

trading partner, irrespective of the roles and learning speeds of the agents. That is, the ToM_2 agent is highly effective in obtaining a larger piece of the pie than his trading partner.

Note that since agents use theory of mind by attributing their own beliefs to their trading partner, a theory of mind agent only has an accurate model of his trading partner's mental content when their learning speeds are the same. For a ToM_1 agent, Figs. 8 and 9 indeed show a high score along the line of equal learning speeds. Interestingly, however, the same is not true for the ToM_2 agents. Figures 11 and 12 show that there is no increased score for the ToM_2 agent along the line of equal learning speeds. That is, a ToM_2 agent benefits from the ability to attribute first-order theory of mind to his trading partner, even if the agent's model of his trading partner's beliefs is inaccurate.

The negotiation scores of two ToM_2 agents negotiating with each other are shown in Fig. 13. Interestingly, the performance of the ToM_2 initiator in Fig. 13a is quite similar to the performance of the ToM_2 responder shown in Fig. 13b. Whereas results from lower orders of theory of mind show many opportunities to divide the pie in one large piece and one small piece, ToM_2 agents generally divide the pie in two pieces that are similar in size. As a result, the graphs in Fig. 13 show little variation in color. Nevertheless, ToM_2 agents that have a positive learning speed that is close to zero tend to do slightly better than ToM_2 agents that have a different learning speed.

In our agent model, a theory of mind agent always starts a negotiation by reasoning at the highest order of theory of mind available to him. However, an agent may choose to play as if he were an agent of a lower order of theory of mind. This decision is based on the agent's confidence c_k in the use of kth-order theory of mind. Analysis of our simulation runs shows that agents typically lose confidence in the use of theory of mind in the beginning of the negotiation. A theory of mind agent believes that the behavior of his trading partner depends on his goal location. However, agents do not know the goal location of their trading partners at the start of a negotiation. As a result, a theory of mind agent loses confidence in the use of theory of mind. Over the course of negotiation, however, the agent obtains information about his trading partner's goal location and regains confidence in the use of theory of mind.

In summary, the results in this section show that the ability to make use of theory of mind can help individuals to negotiate better, although they do not show the same pattern as found for competitive games [13] as predicted by hypothesis H_1 in Sect. 3. Even though the presence of ToM_1 agents prevent negotiation failure in our simulations, the ToM_1 agent does not have a direct competitive advantage over a ToM_0 agent. Instead, the ToM_1 agent suffers the cost for achieving a cooperative solution, which leaves the ToM_0 agent with the larger piece of pie.

The ToM_2 agent, on the other hand, does outperform the ToM_1 agent as expected by hypothesis H_1. Our results show that the ToM_2 agent can negotiate successfully with a ToM_1 trading partner, resulting in a pie that includes a large piece for the ToM_2 agent. When two ToM_2 agents negotiate with each other, the resulting pie is divided into pieces of a fairly similar size. In the next subsection, we take a closer look at the cooperative abilities of these theory of mind agents.

6.2 Social welfare results

In the previous section, we compared the individual competitive performance of agents of various orders of theory of mind negotiating in Colored Trails. In this section, we show how theory of mind affects the cooperative ability of agents, by looking at the social welfare that theory of mind agents achieve through negotiation, where social welfare is measured by the sum of the scores of the initiator and the responder. Figure 14 shows the increase in social welfare for different combinations of theory of mind initiators and responders.

Figure 14a shows that ToM_0 agents can cooperate surprisingly well. In the best-case scenario, both the ToM_0 initiator and the ToM_0 responder have a learning speed of around $\lambda = 0.2$. In this case, negotiators obtain an average social welfare of over 400 points in 4.4 turns of negotiation. However, due to the competitive element of Colored Trails described in Sect. 6.1, cooperation among ToM_0 agents is not stable. The ToM_0 agents experience an evolutionary pressure towards zero learning speed, which can eventually lead to negotiation failure.

Although Sect. 6.1 shows that the presence of a ToM_1 agent can ensure that negotiation failure does not occur, Fig. 14 shows that this does not imply a higher social welfare. Figure 14b, c do not show an improvement over the performance of ToM_0 agents shown in Fig. 14a.

Figure 14d shows that when two ToM_1 agents play Colored Trails together, they achieve the highest social welfare when both agents have a learning speed as high as possible. However, the competitive element in Colored Trails puts an evolutionary pressure on ToM_1 agents to lower their learning speed to a value of $\lambda = 0.2$. Although this does not lead to a breakdown of negotiation like in the case of ToM_0 agents, social welfare suffers from the lower learning speed of ToM_1 agents. The individual desire of ToM_1 agents to obtain as large a piece of pie as possible results in a smaller pie to share.

Although the increase in social welfare depends greatly on the learning speed of ToM_0 agents and ToM_1 agents, Fig. 14e, f show that the learning speed of a ToM_2 agent has little effect on social welfare when a ToM_2 agent and a ToM_1 agent negotiate in Colored Trails. Instead, how negotiation affects social welfare in these cases is determined mostly by the learning speed of the ToM_1 agent.

When two ToM_2 agents negotiate, the highest social welfare is achieved when both agents have a low but positive learning speed, as shown in Fig. 15. Note that this learning speed also yields them the highest score individually. That is, when two ToM_2 agents negotiate, the learning speed that would yield an agent the largest piece of pie is also the learning speed

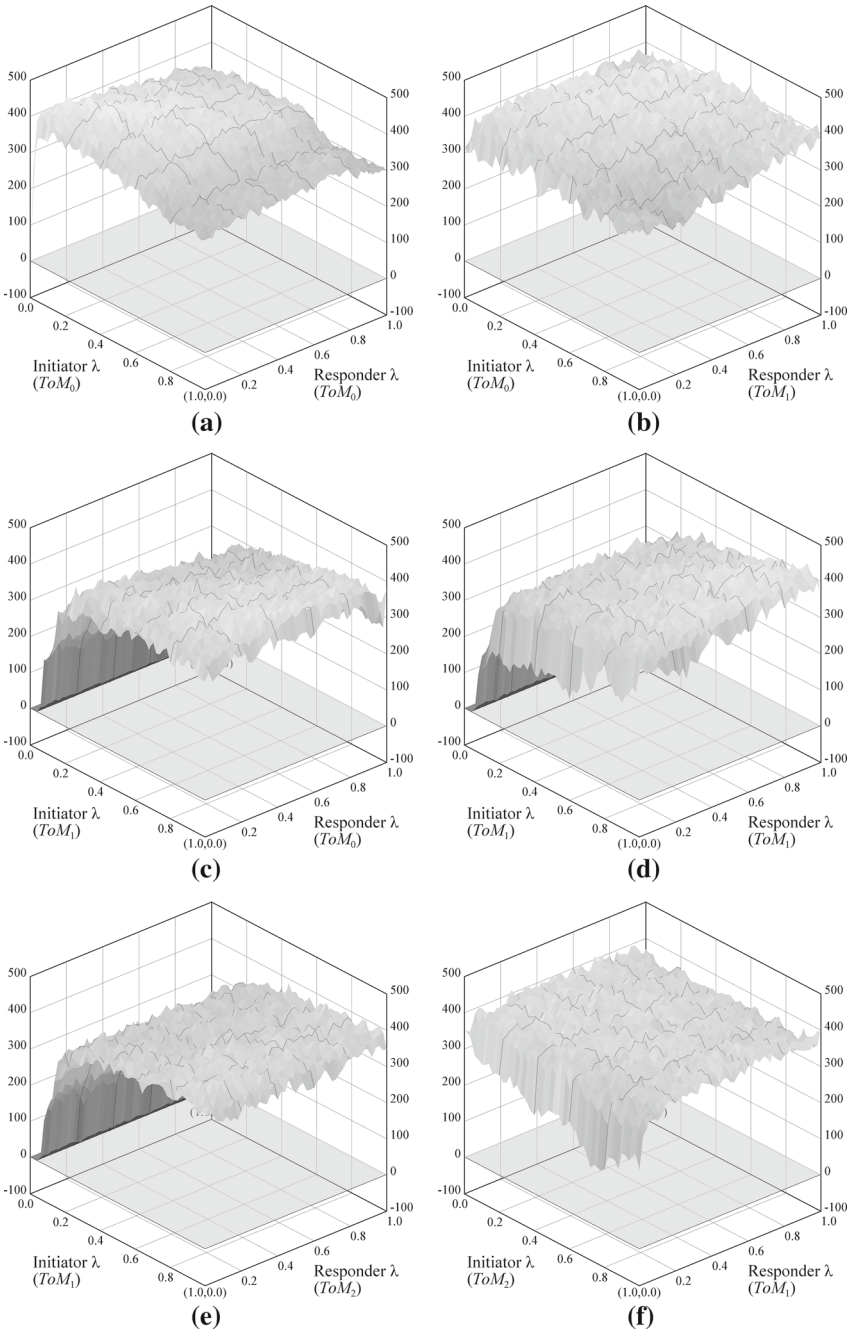

Fig. 14 Average combined negotiation score of theory of mind agents playing Colored Trails. **a** ToM_0 initiator, ToM_0 responder. **b** ToM_0 initiator, ToM_1 responder. **c** ToM_1 initiator, ToM_0 responder. **d** ToM_1 initiator, ToM_1 responder. **e** ToM_1 initiator, ToM_2 responder. **f** ToM_2 initiator, ToM_1 responder

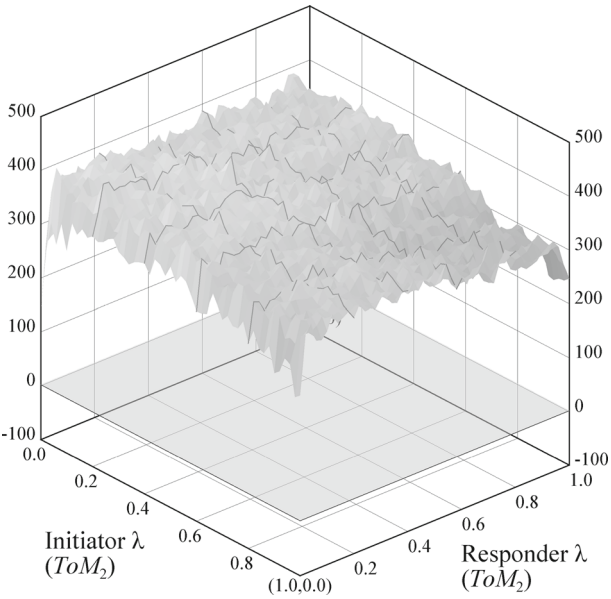

Fig. 15 Average combined negotiation score of two ToM_2 agents playing Colored Trails

that would yield the largest pie to share. However, note that the highest social welfare that the ToM_2 agents achieve is not as high as the highest social welfare achieved by ToM_0 agents. This decrease in social welfare is partially caused by an increase in negotiation length. Especially at the start of a negotiation, a ToM_2 agent makes offers that he expects to be rejected by his trading partner. Rather, the ToM_2 agent expects that making the offer results in a counteroffer from the trading partner that the ToM_2 agent is willing to accept.

In summary, contrary to hypothesis H_2 formulated in Sect. 3, our simulation results do not provide any evidence to support that theory of mind directly increases social welfare. However, we find that theory of mind can help to stabilize negotiation. While ToM_0 agents have the potential to negotiate a high social welfare, natural selection would favor those ToM_0 agents that increase their individual score at the expense of social welfare. These ToM_0 agents therefore face a social dilemma similar to the prisoner's dilemma, which leads ToM_0 agents to a situation in which negotiation breaks down.

A ToM_1 agent is able to avoid complete breakdown of negotiation by following a strategy that also takes the goals of the trading partner into account. However, ToM_1 agents face a similar social dilemma in which the individual desire to obtain as large a piece of pie as possible leads to a smaller pie to share. Interestingly, ToM_2 agents do not face the same social dilemma. When a ToM_2 agent negotiates with a ToM_1 or ToM_2 trading partner, the individual goal to obtain as large a piece of pie as possible leads to a pie for which the total size is as large as possible as well. In this sense, higher-order theory of mind can benefit social welfare in negotiation settings such as Colored Trails.

7 Human participant experiments

In the previous sections, we showed how the use of higher orders of theory of mind can help to stabilize negotiation using simulations with computational agents. Since humans are

known to engage in higher-order theory of mind reasoning, participants may take advantage of the benefits of higher-order theory of mind reasoning when negotiating with others. In this section, we test this prediction in experiments with human participants.[4]

Note that these experiments are not meant to fit the model of computational theory of mind agents to participant data. Rather, by letting participants play against computational agents, we aim to determine to what extent participants make use of theory of mind in these negotiation games. Since our theory of mind agents estimate the level of sophistication of their trading partner, we can use these computational agents to determine to what extent participant behavior is consistent with higher-order theory of mind reasoning. Specifically, by observing the actions of a participant, a ToM_3 agent determines whether it is more likely that the participant is using zero-order, first-order, or second-order theory of mind reasoning. In addition, by varying the level of sophistication of the computational agent, we can show whether participants adjust their level of theory of mind reasoning in response to the behavior of their trading partner.

The remainder of this section is divided as follows. Section 7.1 describes the details of the experimental setup. The results of this analysis are presented in Sect. 7.2.

7.1 Experimental setup

7.1.1 Participants

Twenty-seven students (10 female) of the University of Groningen participated in this study. All participants were informed that after the conclusion of the study, the three participants with the highest score in the negotiation game would receive €15, €10, and €5, respectively. Each participant gave informed consent prior to admission into the study.

7.1.2 Materials

Human participants played a simplified variation of the computational agent simulation experiment of Sect. 6. Instead of the 1,200 negotiations that computational agents played, participants played 24 different negotiation games, each on a different game board. To further simplify the setting for human participants, the 1 point penalty per round of negotiation was removed. Instead, each game was restricted to six rounds of negotiation. That is, once six offers had been made in the same negotiation game, players could no longer choose to make a counteroffer, and were forced to either accept the offer of their trading partner or withdraw from negotiation.

To ensure that these games would allow us to distinguish between different orders of theory of mind reasoning of participants, games were selected for this experiment according to the following conditions:

- The participant's goal could be reached with the eight chips in the game;
- Simulations with computational agents predicted different outcomes for trading partners of different orders of theory of mind; and
- Simulations with computational agents predicted that the game would last at least two turns and at most six turns.

These 24 games were divided into three blocks of eight games each. Each block was associated with a level of theory of mind reasoning of the computational trading partner so that each

[4] The experiment was set up and performed by Eveline Broers.

participant played eight negotiation games with a ToM_0 trading partner, a ToM_1 trading partner, and a ToM_2 trading partner each.

7.1.3 Design and procedure

Because colors played a significant role in the Colored Trails games, participants were tested on colorblindness before the start of the experiment. All participants passed this colorblindness test.[5] The Colored Trails experiment consisted of a familiarization phase and an experimental phase. At the start of the familiarization phase, participants were asked to imagine themselves as an attorney for a major corporation. In this function, they would be involved in a number of negotiations with different clients. Participants were told that their trading partner was Alex, a computer player that would always react on their offer as quickly as possible in a way it believed would maximize its own score. To ensure understanding of the Colored Trails game, participants answered a few questions about the rules, scoring, and movements on the game board. Participants answered these questions correctly.

In the experimental phase, participants played three blocks of eight games each. In each block, the participant negotiated with either a ToM_0, a ToM_1, or a ToM_2 agent. The order in which participants encountered these different trading partners was counterbalanced across participants. Although participants were informed that they would face different clients, they were not told that level of reasoning of the trading partner would change over the course of the experiment. At the start of the experiment, it was randomly decided whether the participant or the computational agent would make the initial offer of the first game. In subsequent games, participant and agent alternated in the role of initiator.

Participants were allowed 60 seconds to decide on their next action. During each round, the remaining decision time was presented to participants by means of a countdown timer. If a participant had not made a decision within 60 seconds, the game continued without an offer being made, and the computational agent took its turn.

The zero-order beliefs of theory of mind agents were initialized by playing 200 randomly generated Colored Trails games against a computational ToM_0 agent. At the start of each game, the agent's beliefs were reset to this initial state. This generic initial state of the computational agent's beliefs allowed us to compare the performance of participants more easily. Additionally, theory of mind agents started every game reasoning at the highest order of theory of mind available to them. This means that although computational agents learned from a participant's offers within a single game and adjusted their behavior accordingly, agents did not exhibit any learning across games. As a result, agents were prevented from adapting to specific participants, and every participant faced the same agent in every scenario. This is consistent with our cover story, in which participants were told that they would negotiate with different clients.

After the Colored Trails games, participants answered a short questionnaire about the perceived difficulty of the task, the behavior of their trading partner, and the participant's reasoning strategies. In addition, participants took a test for their interpersonal reactivity index [11].

[5] Before the start of the Colored Trails experiments, participants also played several Marble Drop games [47]. Participant performance on these Marble Drop games was not correlated with their performance on the Colored Trails games.

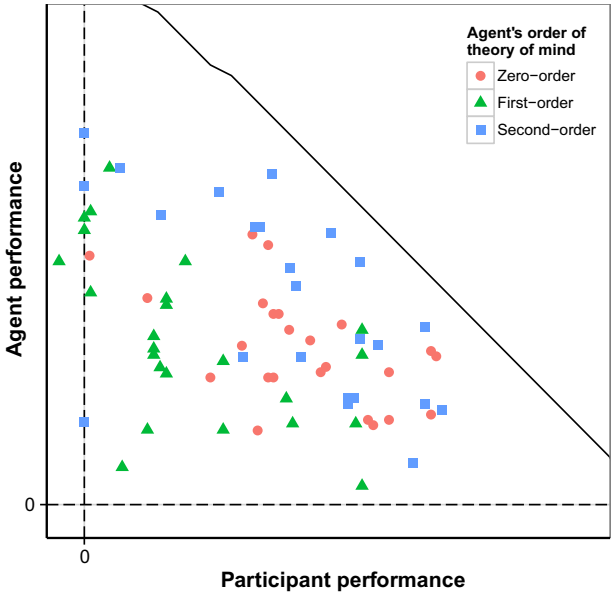

Fig. 16 Increase in score as a result of negotiation in the Colored Trails game for both the human participant and the computational trading partner, when negotiating with a ToM_0 agent (*red circles*), a ToM_1 agent (*green triangle*), and a ToM_2 agent (*blue square*). The solid line indicates Pareto optimal outcomes. *Dashed lines* indicate the score participants and computational agents would achieve if they were to withdraw from negotiation in every game (Color figure online)

7.2 Results

Figure 16 shows the outcomes of the Colored Trails game. For each participant, the figure shows the increase in score as a result of negotiation in the Colored Trails game of both the participant and the computational trading partner, when playing against a ToM_0 agent (red circles), a ToM_1 agent (green triangles), and a ToM_2 agent (blue squares). Dashed lines indicate the zero performance line, which is the score that a player would have received if every game of the block had ended with withdrawal from negotiation. A score below or to the left of the dashed line indicates that a player has decreased his score as a result of negotiation. As Fig. 16 shows, only once a participant received a negative score in one of the blocks. Computational agents never accepted an offer that would have reduced their score.

The solid line in Fig. 16 shows the boundary of Pareto efficient outcomes. This boundary shows those outcomes for which neither the participant nor the computational agent could have received a higher score without a decrease in the score of the other player. Note that since chips are more valuable to a player who can use them to move closer to his goal location, the Pareto boundary shows some discontinuities. The distance of a data point to the Pareto boundary gives an impression of how well participants and computational agents played Colored Trails. Figure 16 shows that participants and computational agents generally negotiated mutually beneficial solutions, while neither player systematically exploited the other. Additionally, on average, the blue squares in Fig. 16 are closer to the Pareto boundary, while green triangles are farther away. This suggests that when participants negotiate with ToM_2 agents, they tend to negotiate better outcomes, while negotiations between participants and ToM_1 agents are typically less successful. The experimental data show that when negotiating with

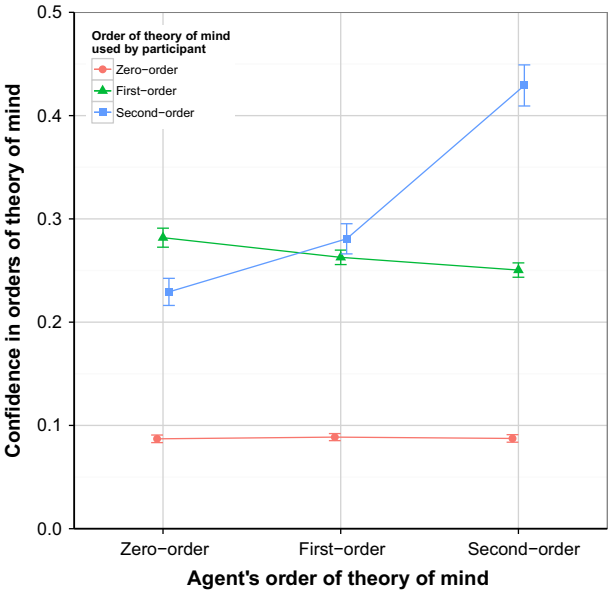

Fig. 17 Estimated similarity of participant offers to the offers of a ToM_0 agent (*red circles*), a ToM_1 agent (*green triangles*), and a ToM_2 agent (*blue squares*) in each of the three blocks, as diagnosed by a ToM_3 agent. *Brackets* indicate one standard error (Color figure online)

ToM_0 agents, participant scores are 60 points higher than agent scores ($W = 430$, $p < 0.01$). However, when paired with a ToM_1 agent, participants score 50 points lower than their trading partner ($W = 165$, $p < 0.02$). Interestingly, the scores of ToM_2 agents are not significantly different to that of their human trading partners ($W = 292.5$, ns).

Our theory of mind agents allow us to estimate to what extent participants make use of theory of mind while playing Colored Trails. We use a ToM_3 'spectator' agent that observes the offers of a participant and adjusts his confidences c_k in kth-order theory of mind accordingly. These confidences c_k give insight in whether the behavior of participants is more indicative of zero-order, first-order, or second-order theory of mind reasoning.

For each of the three blocks, Fig. 17 shows how similar participant offers were to offers of ToM_0, ToM_1, and ToM_2 agents, as judged by the ToM_3 spectator agent. Red circles indicate the average similarity of participant offers to zero-order theory of mind reasoning, green triangles indicate the similarity to first-order theory of mind reasoning, and blue squares show the similarity to second-order theory of mind reasoning. Interestingly, Fig. 17 shows that participant offers are more similar to first-order and second-order theory of mind reasoning than they are to zero-order theory of mind reasoning.

Figure 17 also shows that the similarity ratings of participant offers vary depending on the order of theory of mind of the computer trading partner. Although similarity ratings for zero-order and first-order theory of mind reasoning show no variation across different levels of sophistication of the trading partner ($X^2_{(2)} = 0.52$, ns, and $X^2_{(2)} = 2.67$, ns, respectively), participant offers were significantly more similar to second-order theory of mind reasoning when they were facing a ToM_2 trading partner ($X^2_{(2)} = 24.89$, $p < 0.001$).

In addition to the theory of mind level of the computational agent, the identity of the initiator influenced negotiation outcomes in our setting. The opening bid of a negotiation can serve as an anchor for the entire negotiation process [61,69], making the first offer

particularly influential in the negotiation process. In our experiment, both the participant and the computational agent ended up with an extra 15 points on average after negotiation when the computational agent made the initial offer rather than when the human participant was the first to propose a trade. The only exception to this rule was that participants negotiating with a ToM_2 agent ended up with a higher score when they made the initial offer themselves. This effect can be explained by the way agents of different orders of theory of mind construct their offers. Both ToM_0 and ToM_1 agents make offers that they believe will be accepted by their trading partner. In contrast, ToM_2 agents tend to make offers to change the beliefs and the behavior of their trading partner. As a result, initial offers made by ToM_0 and ToM_1 agents are typically more favorable to their trading partner than those made by ToM_2 agents. Similarly, when participants reasoned more like ToM_2 agents, their initial offers were more favorable to themselves than to their trading partner.

In conclusion, our experiments show that human participants make offers that are more consistent with second-order theory of mind reasoning when their trading partner is capable of second-order theory of mind as well. Interestingly, while participants knew that they would negotiate with a number of different trading partners, they were unaware that these trading partners differed in their theory of mind abilities. That is, the behavior of higher-order theory of mind agents apparently encouraged participants to make use of higher-order theory of mind as well within a few rounds of play. The results of these experiments with human participants confirm that computational agents can benefit from the use of higher-order theory of mind reasoning. More importantly, these results show that human participants can also take advantage of these benefits.

8 Discussion and conclusion

We have investigated the claim that the human ability for higher-order theory of mind may have arisen because of the existence of mixed-motive settings in which the use of higher-order theory of mind presents individuals with an evolutionary advantage [71]. For that purpose, we have simulated interactions between computational agents to show how higher orders of theory of mind can help in obtaining better outcomes in negotiation. In an experiment in which human participants interact with these computational agents, we have also shown that humans indeed take advantage of the benefits of higher-order theory of mind reasoning in negotiations.

We investigated a particular mixed-motive setting known as Colored Trails [28,45,70], which serves as a prototypical multi-issue bargaining situation in which a wide variety of negotiation scenarios can be modeled. In this setting, we let agents of various orders of theory of mind alternate in offering a redistribution of chips under incomplete information about the preferences of their trading partner. In our agent model, a computational agent makes use of theory of mind by taking the position of his trading partner and calculating what his own actions would have been in that position. This approach differs from models of belief hierarchies (e.g., see [35–37]), in which a first-order theory of mind agent defines a probability distribution over all possible zero-order beliefs of his trading partner. Belief hierarchies provide a more general model of theory of mind abilities, but they also assume that a first-order theory of mind agent has an accurate model of any zero-order theory of mind agent. Our approach shows that agents can benefit from the use of higher-order theory of mind, even if they do not have such an accurate model. That is, the ability of theory of mind can emerge even if it does not provide a completely accurate model of the mental content of others.

We found that under the right conditions, agents without any theory of mind abilities could successfully negotiate a mutually beneficial trade in Colored Trails. However, these agents experience an evolutionary pressure to increase their own score at the expense of their trading partner. Through natural selection, this eventually leads to a situation in which all negotiation fails. For zero-order theory of mind agents, negotiation in Colored Trails is similar to the prisoner's dilemma, where the competitive aspect of getting as much of the pie as possible overshadows the cooperative aspect of negotiation to the point where there no longer is any pie to share.

By reasoning explicitly about the goals of the trading partner, first-order theory of mind agents prevent a complete breakdown in negotiation. However, while there are purely competitive settings in which first-order theory of mind agents outperform zero-order theory of mind agents [13,18,75], we find that the same is not true in our mixed-motive setting. The reason for this difference is that in strictly competitive situations, an agent that increases his own score does so by decreasing the score of his opponent. In a mixed-motive situation, an agent that increases his own score may increase the score of his trading partner as well. In Colored Trails, the first-order theory of mind agent increases his own score by preventing negotiation failure. However, this increases the score of a zero-order theory of mind trading partner even more. As a result, the zero-order theory of mind agent obtains a larger piece of the pie. Still, from the perspective of the first-order theory of mind agent, obtaining a small piece of pie is preferable to obtaining no pie at all. In future work, it would be interesting to determine whether the same holds in games with a different balance of cooperation and competition, such as in revelation games [53].

Although first-order theory of mind has a limited effectiveness in the negotiation setting we describe (cf. [59]), second-order theory of mind benefits agents greatly. When a second-order theory of mind agent negotiates with another agent capable of theory of mind, the second-order theory of mind agent typically receives the larger share of the pie. Additionally, neither agent has an incentive to deviate from the outcome that maximizes total pie size. That is, second-order theory of mind provides agents with a strategy that balances cooperative and competitive goals to the point where agents that succeed in negotiating the largest total pie possible could not have received a larger piece of pie for themselves by changing their behavior. Interestingly, this cooperative solution is achieved purely through calculated self-ishness. Second-order theory of mind agents behave cooperatively, not because they have an innate sense of fairness or because they derive utility from the score of their trading partner, but because they believe that it will result in a better outcome for themselves.

Our agent simulation results show that in mixed-motive settings such as negotiations, agents can benefit from the use of higher-order theory of mind. In our experiments with human participants, negotiations with ToM_2 agents indeed resulted in outcomes that were closer to the Pareto optimal boundary. In addition, both players ended up with a higher score on average when the theory of mind agent made the initial offer than when the human participant made the opening bid.

Surprisingly, negotiating with a ToM_2 agent apparently encouraged participants to make use of second-order theory of mind as well. Moreover, participants adjusted their behavior relatively quickly. In the literature, experiments with adults typically show that individuals reason at low orders of theory of mind, and are slow to adjust to an opponent that reasons using theory of mind [8,34,38,75]. However, in our Colored Trails setting, participants exhibited second-order theory of mind within a few games. These results suggest that theory of mind agents can encourage the use of higher-order theory of mind in human participants and may play a useful role in training people to negotiate better outcomes.

Acknowledgments We thank Eveline Broers for designing, implementing, and administering the human participant experiment that we analyze here. This work was supported by the Netherlands Organisation for Scientific Research (NWO) Vici grant NWO 277-80-001, awarded to Rineke Verbrugge for the project 'Cognitive systems in interaction: Logical and computational models of higher-order social cognition'.

References

1. Apperly, I. (2011). *Mindreaders: The cognitive basis of "theory of mind"*. Hove: Psychology Press.
2. Arad, A., & Rubinstein, A. (2012). The 11–20 money request game: A level-k reasoning study. *The American Economic Review*, *102*(7), 3561–3573.
3. Arslan, B., Hohenberger, A., & Verbrugge, R. (2012). The development of second-order social cognition and its relation with complex language understanding and memory. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1290–1295).
4. Bacharach, M., & Stahl, D. O. (2000). Variable-frame level-n theory. *Games and Economic Behavior*, *32*(2), 220–246.
5. Baker, C.L., Saxe, R.R., & Tenenbaum, J.B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2469–2474).
6. Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton, NJ: Princeton University Press.
7. Byrne, R., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford: Oxford University Press.
8. Camerer, C., Ho, T., & Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*(3), 861–898.
9. Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874.
10. Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*(5), 1193–1235.
11. Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126.
12. de Weerd, H., & Verbrugge, R. (2011). Evolution of altruistic punishment in heterogeneous populations. *Journal of Theoretical Biology*, *290*, 88–103. doi:10.1016/j.jtbi.2011.08.034.
13. de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, *199–200*, 67–92. doi:10.1016/j.artint.2013.05.004.
14. de Weerd, H., Verbrugge, R., & Verheij, B. (2015). Higher-order theory of mind in the tacit communication game. *Biologically Inspired Cognitive Architectures*, *11*, 10–21. doi:10.1016/j.bica.2014.11.010.
15. de Jong, S., Hennes, D., Tuyls, K., & Gal, Y. (2011). Metastrategies in the colored trails game. In *Proceedings of 10th international conference on autonomous agents and multiagent systems (IFAAMAS)* (pp. 551–558).
16. de Weerd, H., Verbrugge, R., & Verheij, B. (2014). Agent-basedmodels for higher-order theory of mind. In *Advances in social simulation, proceedings of the 9th conference of the european socialsimulation association* (vol. 229, pp. 213–224). Berlin: Springer.
17. Devaine, M., Hollard, G., & Daunizeau, J. (2014). The social Bayesian brain: Does mentalizing make a difference when we learn? *PLoS Computational Biology*, *10*(12), e1003,992. doi:10.1371/journal.pcbi.1003992.
18. Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of mind: Did evolution fool us? *PloS One*, *9*(2), e87,619. doi:10.1371/journal.pone.0087619.
19. Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2733–2742.
20. Doshi, P., Qu, X., Goodie, A., & Young, D. (2010). Modeling recursive reasoning by humans using empirically informed interactive POMDPs. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems (IFAAMAS)* (vol. 1, pp. 1223–1230).

21. Dunbar, R. (1998). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.
22. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.
23. Fatima, S., Kraus, S., & Wooldridge, M. (2014). *Principles of automated negotiation*. Cambridge: Cambridge University Press.
24. Ficici, S.G., & Pfeffer, A. (2008). Modeling how humans reason about others with partial information. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems (IFAAMAS)* (pp. 315–322).
25. Fisher, R., & Ury, W. L. (1981). *Getting to yes: Negotiating agreement without giving in*. London: Penguin Group.
26. Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, *17*(4), 417–442.
27. Franke, M., & Galeazzi, P. (2014). On the evolution of choice principles. In Szymanik, J., & Verbrugge, R. (eds.) *Proceedings of the second workshop reasoning about other minds: Logical and cognitive perspectives, co-located with advances in modal logic, groningen, CEUR workshop proceedings* (vol. 1208, pp. 11–15).
28. Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., & Shieber, S. (2010). Agent decision-making in open mixed networks. *Artificial Intelligence*, *174*(18), 1460–1480.
29. Gintis, H. (2009). *The bounds of reason: Game theory and the unification of the behavioral sciences*. Princeton, NJ: Princeton University Press.
30. Gmytrasiewicz, P., & Durfee, E. (1995). A rigorous, operational formalization of recursive modeling. In *Proceedings of the first international conference on autonomous agents and multiagent systems* (pp. 125–132).
31. Gmytrasiewicz, P. J., & Doshi, P. (2005). A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, *24*, 49–79.
32. Gmytrasiewicz, P. J., Noh, S., & Kellogg, T. (1998). Bayesian update of recursive agent models. *User Modeling and User-Adapted Interaction*, *8*(1–2), 49–69.
33. Goeree, J. K., & Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior*, *46*(2), 365–382.
34. Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, *25*(1), 95–108.
35. Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players part I. The basic model. *Management Science*, *14*(3), 159–182.
36. Harsanyi, J. C. (1968). Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points. *Management Science*, *14*(5), 320–334.
37. Harsanyi, J. C. (1968). Games with incomplete information played by "Bayesian" players, part III. The basic probability distribution of the game. *Management Science*, *14*(7), 486–502.
38. Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*(1), 1–36.
39. Heyes, C. (2012). Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703.
40. Hu, J., & Wellman, M.P. (1998). Online learning about other agents in a dynamic multiagent system. In *Proceedings of the second international conference on autonomous agents (ACM)* (pp. 239–246).
41. Kraus, S. (1997). Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, *94*(1), 79–97.
42. Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*(2), 141–160. doi:10.1016/j.cogsys.2006.07.004.
43. Lin, R., Gal, Y. K., Kraus, S., & Mazliah, Y. (2014). Training with automated agents improves people's behavior in negotiation and coordination tasks. *Decision Support Systems*, *60*, 1–9.
44. Lin, R., Kraus, S., Baarslag, T., Tykhonov, D., Hindriks, K., & Jonker, C. M. (2014). Genius: An integrated environment for supporting the design of generic automated negotiators. *Computational Intelligence*, *30*(1), 48–70.
45. Lin, R., Kraus, S., Wilkenfeld, J., & Barry, J. (2008). Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence*, *172*(6), 823–851.
46. McKelvey, R., & Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, *10*(1), 6–38.
47. Meijering, B., van Rijn, H., Taatgen, N., & Verbrugge, R. (2011). I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In *Proceedings of the 33nd annual conference of the cognitive science society* (pp. 2486–2491).

48. Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773.
49. Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 639–648.
50. Nagel, R. (1995). Unraveling in guessing games: An experimental study. The American Economic Review, 1313–1326.
51. Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563.
52. Pacuit, E. (2015). Dynamic logic and strategic reasoning. In J. van Benthem, S. Ghosh, & R. Verbrugge (Eds.), *Modeling strategic reasoning: Logics, games, and communities. Lecture notes in computer science* (Vol. 7081). Heidelberg: Springer.
53. Peled, N., Gal, Y., & Kraus, S. (2015). A study of computational and human strategies in revelation games. *Autonomous Agents and Multi-Agent Systems*, *29*(1), 73–97.
54. Penn, D., & Povinelli, D. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 731–744.
55. Perea, A. (2012). *Epistemic game theory: Reasoning and choice*. Cambridge: Cambridge University Press.
56. Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that.." attribution of second-order beliefs by 5 to 10 year old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. doi:10.1016/0022-0965(85)90051-7.
57. Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.
58. Pynadath, D.V., & Marsella, S.C. (2005). PsychSim: Modeling theory of mind with decision-theoretic agents. In IJCAI (pp. 1181–1186).
59. Pynadath, D. V., Rosenbloom, P. S., & Marsella, S. C. (2014). Reinforcement learning for adaptive theory of mind in the Sigma cognitive architecture. In B. Goertzel, L. Orseau, & J. Snaider (Eds.), *Artificial general intelligence. Lecture notes in artificial intelligence* (Vol. 8598, pp. 143–154). Berlin: Springer.
60. Qu, X., Doshi, P., & Goodie, A. (2012). Modeling deep strategic reasoning by humans in competitive games. In van der Hoek, W., Padgham, L., Conitzer, V., & Winikoff, M. (eds.) *Proceedings of the 11th international conference on autonomous agents and multiagent systems (IFAAMAS)* (vol. 3, pp. 1243–1244).
61. Raiffa, H., Richardson, J., & Metcalfe, D. (2002). *Negotiation analysis: The science and art of collaborative decision making*. Cambridge: Belknap Press.
62. Rosenfeld, A., Zuckerman, I., Segal-Halevi, E., Drein, O., & Kraus, S. (2014). NegoChat: A chat-based negotiation agent. In Lomuscio, A., Scerri, P., Bazzan, A., & Huhns, M. (eds.) *Proceedings of the thirteenth international conference on autonomous agents and multi-agent systems* (pp. 525–532).
63. Rosette, A. S., Kopelman, S., & Abbott, J. L. (2013). Good grief! Anxiety sours the economic benefits of first offers. *Group Decision and Negotiation*, *23*, 1–19.
64. Stahl, D., & Wilson, P. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*(1), 218–254.
65. Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.
66. van der Post, D. J., de Weerd, H., Verbrugge, R., & Hemelrijk, C. K. (2013). A novel mechanism for a survival advantage of vigilant individuals in groups. *The American Naturalist*, *182*(5), 682–688.
67. van der Vaart, E., Verbrugge, R., & Hemelrijk, C. (2012). Corvid re-caching without 'theory of mind': A model. *PLoS One*, *7*(3), e32,904.
68. van Ditmarsch, H., van der Hoek, W., & Kooi, B. P. (2007). *Dynamic epistemic logic*. Berlin: Springer.
69. van Poucke, D., & Buelens, M. (2002). Predicting the outcome of a two-party price negotiation: Contribution of reservation price, aspiration price and opening offer. *Journal of Economic Psychology*, *23*(1), 67–76.
70. van Wissen, A., Gal, Y., Kamphorst, B., & Dignum, M. (2012). Human-agent teamwork in dynamic environments. *Computers in Human Behavior*, *28*(1), 23–33.
71. Verbrugge, R. (2009). Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic*, *38*, 649–680.
72. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
73. Whiten, A., & Byrne, R. (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge: Cambridge University Press.
74. Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

75. Wright, J.R., & Leyton-Brown, K. (2010). Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the twenty-fourth conference on artificial intelligence* (pp. 901–907).
76. Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology*, *4*(12), e1000,254.
77. Zhang, J., Hedden, T., & Chia, A. (2012). Perspective-taking and depth of theory-of-mind reasoning in sequential-move games. *Cognitive Science*, *36*(3), 560–573.